Hola UniVRso: Realidad virtual para oradores

Trabajo de Fin de Máster

Curso 2017–2018

Autora
Meriem El Yamri El Khatibi

Director
Borja Manero Iglesias

Colaboradora
Eva Ullán

Máster en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

Hola UniVRso: Realidad virtual para oradores

Trabajo de Fin de Máster en Ingeniería Informática
Departamento de Ingeniería del Software e Inteligencia Artificial

Autora
Meriem El Yamri El Khatibi

Director
Borja Manero Iglesias

Colaboradora
Eva Ullán

Convocatoria: *Junio 2018*
Calificación: *10 (Matrícula de Honor)*

Máster en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

17 de julio de 2018

Autorización de difusión

La abajo firmante, matriculada en el Máster en Ingeniería en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Máster: "Hola UniVRso: Realidad virtual para oradores", realizado durante el curso académico 2017/18 bajo la dirección de Borja Manero Iglesias y Eva Ullán en el Departamento de Ingeniería del Software e Inteligencia Artificial, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Meriem El Yamri El Khatibi

17 de julio de 2018

Hay dos tipos de oradores: los que se ponen nerviosos y los mentirosos

Mark Twain

Agradecimientos

A mis padres, porque a pesar de que voy a visitarles sólo me ven teclear en frente de una pantalla sin saber muy bien qué estoy haciendo.

A mi hermana, por escucharme y asentir.

A Eva, por ser una fuente inagotable de información, por organizar, por encontrar soluciones a cualquier problema, por los atracones de trabajo, y, sobre todo, por confiar en mi idea desde el primer momento.

A Borja, por creer en este proyecto y lanzarte de cabeza a ello, por hablarle sobre este trabajo a todo el mundo, y por no haber rechazado ni una sola de las ideas que se me han ido ocurriendo.

A Rodri, por estar ahí siempre, llueva o truene.

A Jesper Kondrup, por cederme tu modelo de cine para el primer nivel de este proyecto.

A Pablo Gervás y todos los investigadores del proyecto ComunicArte, por aportarme un sinfín de ideas y guiarme en la dirección adecuada.

A la Fundación BBVA por la financiación para el proyecto ComunicArte.

A Niko Verona y Carmen Ibeas, por ayudarme con vuestras interpretaciones en el experimento.

A Marco Antonio y Pedro Pablo Gómez Martín, por haber creado una plantilla de LaTeX que me ha ahorrado tiempo y disgustos dando formato a este documento, y a Gonzalo Méndez por retocarla para TFMs.

Resumen

La oratoria o el arte de hablar en público con elocuencia se ha cultivado desde la antigüedad y ha sido una constante para el ser humano. Los griegos fueron los primeros en establecer una serie de normas para esta disciplina transversal, y posteriormente, hablar en público ha sido la forma en la que muchas personalidades han destacado y lo que les ha permitido cambiar el curso de la historia..

Además, esta disciplina no está sólo al servicio de los grandes líderes, sino que es necesaria en muchas situaciones de la vida cotidiana, desde dar una conferencia hasta al hacer una entrevista de trabajo.

Sin embargo, hablar en público ha constituido siempre un obstáculo manifestado en forma de miedo, que afecta a una parte generalizada de la población, y que consiste en una reacción desproporcionada a la situación de amenaza de enfrentarse a una audiencia.

Por este motivo, este trabajo nace ante la necesidad de ayudar a paliar este miedo y de proporcionar una herramienta de entrenamiento de la habilidad de hablar en público. Para ello, se ha construido un sistema de realidad virtual que ofrece al orador un entorno seguro, similar al mundo real, en el que practicar sus presentaciones. Este sistema se ha desarrollado como un videojuego en el que el ponente se enfrenta a una audiencia virtual reactiva que, basándose en las emociones extraídas de algunos parámetros del orador (i.e. tono de voz y mirada), va reaccionando en tiempo real a su discurso. De esta forma, el orador puede ir adaptando su discurso a la audiencia y recibir realimentación en tiempo real sobre la efectividad de su exposición.

Para modelar el algoritmo de reacciones de la audiencia virtual, y comprobar la efectividad del sistema en conjunto, se han realizado experimentos con un grupo de individuos. Estos experimentos han generado resultados muy positivos, validando la utilidad de este trabajo como herramienta para el entrenamiento de la disciplina de hablar en público, y aportando muchas vías de desarrollo futuro y mejoras a este proyecto.

Además de lo anterior, este trabajo ha sido el germen de un proyecto de investigación más grande, denominado "ComunicArte", que cuenta con financiación para implementar lo que se propone en este documento con un espectro más amplio de funcionalidades y alcance.

Palabras clave

hablar en público, videojuego educativo, realidad virtual, VRET, oratoria, análisis de emociones, retórica

Abstract

Oratory or the art of public speaking eloquently has been cultivated since ancient times, and it has been a constant for human kind. The Greeks were the first people to establish a set of rules for this transversal discipline, and after that, public speaking has been one of the ways in which great personalities have changed the course of history through their oratory abilities.

Furthermore, this discipline is not intended as a tool only for great leaders, but it is also needed in many daily situations: giving a speech or talking to HR in a job interview.

However, public speaking has been an obstacle for a great part of the world's population, manifested in the form of fear, which is an overreaction to the threatening situation of facing an audience.

Due to this matter, the goal of this work was born of the need to try to palliate this fear and offer a training tool to practice the ability of public speaking. For this purpose, a virtual reality system has been built, which offers the orator a safe environment, similar to the real word, where to practice its speeches. The mentioned system has been developed in the form of a videogame where the orator faces a reactive virtual audience. This audience reacts in real time to the emotions transmitted in the orator's speech, which are extracted from certain speaker's parameters (i.e. voice tone and look focus). Thus, the orator can adapt its speech according to the audience reactions and receive feedback of its speech performance.

In this work there have been some experiments with a group of individuals to develop the reactions algorithm and test the whole system effectiveness. These experiments have generated very positive results, validating the the usefulness of this work as a training tool for public speaking, and also contributing to new development paths and project enhancements.

In addition to the above, this work has been the seed of a greater research project, called "ComunicArte", which has been granted financing to

develop what is discussed in this document but with a wider scope and more functionalities.

Keywords

public speaking, educational videogame, virtual reality, VRET, oratory, emotion analysis, rhetoric

Índice

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

Desde la antigüedad, la necesidad de hablar en público ha sido una constante para el ser humano. Hablar delante de un público o grupo de personas se ha utilizado a lo largo de la historia para persuadir, convencer, enseñar o incluso dirigir el pensamiento.

Los griegos, siendo pioneros los sofistas como Protágoras [48], y posteriormente grandes oradores como Demóstenes o Sócrates [60], dieron origen a la **retórica**, un conjunto de disciplinas y una forma concreta de organizar y construir un discurso con un objetivo final: persuadir a la audiencia. La retórica es transversal a distintos campos del conocimiento (ciencia de la literatura, ciencia política, publicidad, periodismo, ciencias de la educación, ciencias sociales, derecho, estudios bíblicos, etc.) y se ocupa de estudiar y de sistematizar procedimientos y técnicas de utilización del lenguaje, puestos al servicio de una finalidad persuasiva o estética, añadida a su finalidad comunicativa.

Posteriormente, los romanos, entre ellos Cicerón [7] y Marco Fabio Quintiliano [44], siguieron perfeccionando esta disciplina con sus publicaciones al respecto.

Acercándonos a nuestra época, numerosas personalidades han destacado por su uso de la oratoria para dirigir el curso de los acontecimientos: entre ellos Abraham Lincoln con su discurso de Gettysburg [8], Martin Luther King con su *I have a dream* [27], Winston Churchill, Nelson Mandela o Steve Jobs, entre otros.

La **oratoria** y el arte de hablar en público con elocuencia, con la finalidad de persuadir o conmover al auditorio, es una habilidad transversal utilizada en muchos ámbitos y que ha sido importante para el ser humano a lo largo

de la historia. En la comunicación oral también se incluye aquello que no se dice (expresión corporal, ritmo, tono de la voz, etc.).

Se trata, por tanto, de una disciplina transversal a diferentes ámbitos de la vida del ser humano: aparte de impartir una conferencia ante un auditorio, también puede ponerse en práctica para hablar en una reunión de vecinos, intervenir en clase, dar un punto de vista y defenderlo, hablar delante de un jefe de recursos humanos en una entrevista de trabajo o incluso dar un discurso en una boda.

Sin embargo, el cultivo de esta disciplina se ha enfrentado a un obstáculo que también ha acompañado al hombre a lo largo de la historia: el miedo a hablar en público.

Según la Sociedad Española para el Estudio de la Ansiedad y el Estrés (SEAS) [39], al hablar en público, la mayor parte de las personas reaccionan con niveles altos de activación, por el esfuerzo que supone el manejo cognitivo de la información (recuperación de la información, relacionar unos elementos con otros, etc.), la verbalización de los contenidos, con un volumen de voz incrementado, manteniendo un alto gasto de recursos (energéticos, atencionales, etc), durante un tiempo prolongado. Este conjunto de factores es lo que se experimenta como nervios o **miedo escénico**. El miedo escénico no es nuevo para aquellos familiarizados con las artes escénicas. Sin embargo, en el teatro, por ejemplo, los actores pueden refugiarse detrás del personaje que encarnan para paliar sus efectos. Pero estas técnicas no resultan apropiadas para dar charlas, ya que una de las premisas del buen orador es "ser uno mismo".

Un porcentaje significativo de personas se activan demasiado, no pudiendo evitar centrar casi toda su atención en sus síntomas de ansiedad y, en consecuencia, hablando en público. Esto hace que la situación resulte en una experiencia tremendamente desagradable, por lo que tratan de evitarla a toda costa. Algunas de las personas que practican la evitación acaban recibiendo un diagnóstico de trastorno de ansiedad por **fobia social**.

Volviendo a lo anterior, según la RAE[1], miedo es:

1. m. Angustia por un riesgo o daño real o imaginario.
2. m. Recelo o aprensión que alguien tiene de que le suceda algo contrario a lo que desea.

y escénico se refiere a:

[1] `www.rae.es`

1. adj. Perteneciente o relativo a la escena.

Se deduce, pues, que el miedo escénico es una angustia porque suceda un acontecimiento (real o imaginario) cuando uno está en un escenario. Existen muchas formas de denominar a este miedo: pánico escénico, miedo a hablar en público, etc. En ocasiones, este miedo escénico deriva en una fobia. De nuevo, según la RAE, fobia es:

1. f. Aversión exagerada a alguien o a algo.
2. m. Psiquiatr. Temor angustioso e incontrolable ante ciertos actos, ideas, objetos o situaciones, que se sabe absurdo y se aproxima a la obsesión.

El miedo escénico contempla otras actuaciones en público además de hablar, ya sea cantar, actuar, bailar, etc. Sin embargo, este documento se centra solamente en el miedo a hablar en público, siendo éste más específico que el miedo escénico.

El nombre técnico para la fobia a hablar en público es **glosofobia** (ansiedad para hablar). Esta alteración se diferencia de la fobia social en el sentido de que en esta última la persona teme cualquier tipo de actividad que requiera socializarse, mientras que en la glosofobia el elemento temido es únicamente la actividad de hablar en público [17]. El temor que siente una persona con glosofobia aparece siempre que dicha persona se enfrente a la actividad de hablar en público y es independiente del contexto y la situación particular. Para este tipo de fobia no es suficiente con aplicar técnicas de exposición, sino que además es necesaria la ayuda profesional.

Sin embargo, el miedo a hablar en público, que nace de la timidez o simplemente de la ansiedad que produce la situación de enfrentarse a una audiencia, está tan extendido entre la población que existe el test de Trier, una prueba creada en los años noventa para medir el estrés social.

> Esta prueba explota precisamente el miedo a hablar en público y consiste en que una serie de voluntarios son conducidos a una habitación donde tres personas les comunican que tienen diez minutos para preparar una presentación de cinco minutos. Hay una cámara que va a grabar la presentación y tras ello los sujetos deben contar desde el número 1022 hasta el 13 hacia atrás, en voz alta y sin cometer fallos.

Este tipo de experimento pone al sujeto en una situación de anticipación al peligro, de forma que se produce una reacción, generalmente intensa, que

surge como consecuencia de **pensamientos anticipatorios catastróficos** sobre la situación real o imaginaria de hablar en público.

El miedo a hablar en público se puede definir como una respuesta desproporcionada del cuerpo a una situación, confundiendo la preocupación y los nervios con una amenaza. De esta forma, el ritmo cardíaco se acelera y puede ir acompañado de otros síntomas como sudores, temblores, falta de aire, enrojecimiento de la piel, pérdida del hilo narrativo e incluso mareos. **El cuerpo se anticipa de esta forma a una situación de amenaza que no es real**.

Por lo tanto, siendo la oratoria una habilidad necesaria en el día a día de las personas y sabiendo que el miedo a hablar en público es una afección padecida por alrededor del 75 % de la población [52], es natural pensar en buscar alguna forma de paliar el problema [50].

1.1. Motivación del trabajo

En la actualidad, la necesidad de hablar en público se da en diferentes ámbitos de la vida del ser humano, desde realizar una presentación en un evento hasta hablar en una entrevista de trabajo. De hecho, en el Programa Marco Horizonte 2020[2], que integra parte de las actividades de La Unión Europea de investigación e innovación en el período 2014-2020, enseñar a los europeos a hablar en público es uno de los apartados que ha tenido gran importancia.

Hablar en público de forma elocuente y conseguir convencer, conmover o persuadir al público es una habilidad transversal de la que hay que hacer uso en diversas situaciones, como las anteriormente mencionadas.

Sin embargo, el miedo escénico y el miedo a hablar en público son de los miedos más prevalentes en el mundo y afectan a una gran parte de la población. Constituyen, por tanto, un obstáculo importante para desarrollar esta habilidad.

A lo largo de la historia, estos miedos se han tratado de diferentes formas: teatro o actividades de improvisación, terapias conductuales, talleres para hablar en público, etc. Y a partir de los años 90 empiezan a proliferar herramientas que hacen uso de la tecnología para abordar miedos o entrenar determinadas habilidades.

Por lo tanto, con la tecnología de la que se dispone en la actualidad, es factible pensar en un sistema que permita: superar o disminuir el miedo escénico; o entrenar la habilidad transversal de hablar en público.

Tal y como se puede ver en el siguiente capítulo, todas las aplicaciones con el mismo propósito que este trabajo, afrontan el reto con mayor o menor éxito, pero en ninguna de ellas se dan conjuntamente los dos elementos que en este trabajo se consideran cruciales: **ofrecer una realimentación al orador y hacer que mejore sus habilidades de presentación mientras la está haciendo**.

En primer lugar, tal y como veremos, solamente dos de las herramientas investigadas incorporan sensores para medir algunas de las características del orador: su discurso y gestos, su volumen y tono de voz, hacia dónde mira o su postura.

En segundo lugar, y más importante, ninguna de las herramientas de la literatura revisada ofrece una audiencia reactiva en tiempo real, basándose

[2] `https://eshorizonte2020.es/`

en los datos obtenidos de dichos sensores.

Por ese motivo, la idea de este trabajo es combinar realidad virtual y sensores que midan las acciones del orador, permitiendo al sistema reaccionar y ofrecer realimentación en tiempo real, para que el ponente pueda modificar su discurso de forma acorde. Además, y en contraposición a las aplicaciones presentadas, lo que aquí se quiere ofrecer al usuario es una **gamificación** del entorno, un modo de juego, y no tanto una terapia o simulador al uso, de forma que la experiencia sea inmersiva y favorezca la motivación por mejorar.

1.1.1. Realidad Virtual y VRET

La **realidad virtual** es un entorno realista de tres dimensiones, compuesto de imágenes o modelos 3D, que se crea mediante el uso de hardware y software interactivo y que se presenta al usuario de forma que éste lo pueda aceptar como un entorno real en el que se interactúa de una manera similar o igual al mundo físico.

En los últimos años, la realidad virtual ha experimentado un gran desarrollo gracias a la creciente disponibilidad para el gran público de dispositivos de realidad virtual.

Hay varios estudios [50, 41] que avalan el uso de la realidad virtual en el contexto del tratamiento de fobias y miedos que experimentan los individuos. La premisa que se debe cumplir en este tipo de tratamientos es la llamada respuesta de **presencia** [49, 3, 16], que es una forma de medir la capacidad de una experiencia interactiva en realidad virtual para hacer que los individuos se sientan "realmente" en un entorno virtual, es decir, que el grado de ansiedad experimentado por el sujeto dentro del entorno de realidad virtual debe ser lo suficientemente similar a cómo reaccionaría de forma normal en una situación similar del mundo real.

Las formas más comunes de presencia son la telepresencia, que es la capacidad de un entorno virtual para crear una ilusión de "estar allí", y la presencia social, que define el sentimiento de "estar allí con otro" y analiza, por ejemplo, las interacciones entre lo real y lo virtual.

Para este proyecto es interesante explorar ambos aspectos de la presencia, ya que hablar en público es una actividad social por definición.

La **inmersión** es la percepción de estar físicamente dentro de un mundo no físico. Engloba el sentido de la presencia, que es el punto en el que el cerebro humano cree que está en un sitio en el que no está realmente, lo cual

se consigue mediante medios puramente mentales y/o físicos. El estado de inmersión total existe cuando hay suficientes sentidos activados para crear la percepción de estar presente en un mundo virtual. Existen dos tipos comunes de inmersión:

1. Inmersión mental: Un estado mental profundo en el que se suspende la creencia de que uno está en un entorno virtual.
2. Inmersión física: Estado en el que ocurre interacción física con el entorno virtual.

El objetivo de la realidad virtual es generar inmersión total, es decir, generar una experiencia sensorial que parezca tan real que el usuario olvide que está dentro de un entorno virtual creado artificialmente, e interactúe con él como lo haría en el mundo real. En un entorno de realidad virtual el mundo es completamente sintético y puede o no imitar las propiedades de un entorno real. Esto implica que la realidad virtual puede simular una situación cotidiana (e.g. andar por las calles de una ciudad, interactuar con un entorno de una oficina, etc.), o puede exceder los límites de lo físico creando un mundo en el que las leyes de la física como la gravedad, el tiempo y las propiedades de los materiales no se sostengan (e.g. disparar a alienígenas en un planeta sin gravedad).

La realidad virtual requiere que se estimulen tantos de nuestros sentidos como sea posible. La estimulación correcta de estos sentidos requiere realimentación sensorial, que se consigue mediante el hardware y software integrados. Ejemplos de este hardware pueden ser los HMD (*head-mounted display*), guantes especiales con realimentación táctil o accesorios de mano o de control [46].

Para que el cerebro humano acepte un entorno artificial y virtual como real, no sólo tiene que parecer este mundo real sino que también ha de "sentirse" real. Que el mundo parezca real se puede conseguir mediante el HMD, que recrea un entorno virtual 3D a tamaño real sin las limitaciones de una pantalla. Para que el mundo se sienta real se pueden usar entradas al sistema como seguidores de movimiento que basan la interactividad en los movimientos del usuario. Si se estimulan todos los sentidos que se utilizan normalmente para funcionar en el mundo real, los entornos de realidad virtual comienzan a dar la misma sensación que el mundo real.

Una de las aplicaciones principales de la realidad virtual está en el mundo de los videojuegos, debido a la capacidad inmersiva de la realidad virtual mencionada anteriormente. Esta aplicación está muy extendida en los diferentes tipos de videojuegos existentes, desde juegos de aventuras en primera

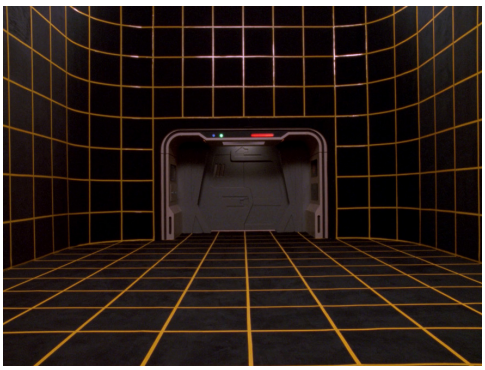

Figura 1.1: Star Trek: Holodeck

persona hasta aventuras gráficas con narrativa de ficción.

Pero este concepto no es, ni mucho menos nuevo. De hecho, ya en 1997 Janet Murray [36] se postuló como defensora de la idea de que algún día los ordenadores podrían ofrecer experiencias interactivas tan realistas que llegaran a ser indistinguibles de la realidad. Esta autora imaginaba un futuro en el que un ordenador pudiera ofrecer narrativas perfectas dentro del Holodeck[3], un dispositivo ficticio consistente en una habitación en la que entraban los personajes de Star Trek para relajarse. En esta habitación, el ordenador creaba un entorno de realidad virtual en el que los participantes se convertían en los protagonistas de una novela. Dependiendo de las interacciones que tuvieran con los personajes de la historia, el ordenador iba creando la continuación de la novela. Para Murray, convertirse en un personaje de una ficción dentro del Holodeck sería una experiencia tanto placentera como de aprendizaje. Al igual que en un videojuego, el Holodeck ofrece un lugar seguro en el que enfrentarse a sentimientos molestos que de otra forma se suprimen; permite reconocer las fantasías más aterradoras sin quedar paralizados por ellas [25].

Sin embargo, los videojuegos no son la única aplicación de la realidad virtual. Existen muchos ámbitos en los que se hace uso de esta tecnología. Por ejemplo, en el ámbito militar se utiliza la realidad virtual para simuladores de vuelo o de guerra, y en el área de salud se utiliza la realidad virtual para simular entornos de operaciones médicas. Hay muchos más casos, pero estos dos ejemplos demuestran la capacidad de la realidad virtual para **generar entornos seguros en los que practicar unas determinadas habilidades**, ya sea pilotar un avión o un tanque, hacer un trasplante de corazón, reaccionar en caso de fuga en una central nuclear o incluso, como plantea este trabajo, practicar la habilidad de hablar en público. Básicamente, el

[3] `http://www.startrek.com/database_article/holodeck`

concepto que aquí se está manejando es el de crear un entorno seguro en el que se pueda fallar sin que esto tenga consecuencias en el mundo real. Un concepto fundamental para que se pueda producir aprendizaje.

Independientemente del tipo, las experiencias de realidad virtual se deben diseñar de forma diferente que en un videojuego al uso, dado que los movimientos de la cabeza en el mundo virtual pueden producir un efecto de mareo conocido como **enfermedad del simulador** o *simulator sickness*. Este efecto fue documentado por primera vez en un simulador de helicóptero por Havron y Butler [19] en 1957 y se refiere al malestar generado por el cuerpo cuando experimenta un movimiento virtual que no tiene lugar en el mundo real.

VRET (del inglés *Virtual Reality Exposure Therapy*, o terapia de choque mediante realidad virtual) [53] consiste en utilizar la realidad virtual en terapia ocupacional o psicológica. En este tipo de terapias los pacientes navegan por un mundo de realidad virtual, creado de forma digital y diseñado específicamente para realizar tareas que permitan tratar una patología específica.

En VRET un individuo se introduce en un entorno virtual generado por ordenador, ya sea usando un HMD o entrando en una habitación con pantallas alrededor. Este entorno se puede programar de forma que la persona se pueda enfrentar a situaciones o lugares que le den miedo y que pueden no ser seguros de confrontar en el mundo real.

Hay algunos estudios [59, 45] que demuestran que la VRET puede ser útil para tratar diferentes trastornos de ansiedad o problemas relacionados con la ansiedad, incluyendo la claustrofobia, el miedo a conducir, el miedo a las alturas, el miedo a volar, el miedo a las arañas o la fobia social. También existen estudios [45] que utilizan VRET para el tratamiento de trastornos de estrés postraumático.

1.1.2. Análisis de emociones

Además del entorno virtual, es importante realizar un análisis de las emociones del orador, para determinar si está experimentando esa sensación de presencia y para calibrar las reacciones de la audiencia virtual.

A pesar de que en el área de VRET hay bastante investigación con respecto a su uso para tratar el miedo a hablar en público, en el ámbito de modelar una audiencia virtual no se ha conseguido encontrar ningún trabajo que haga un análisis de parámetros del orador (hacia dónde mira, tono y

proyección de voz, gestos, etc.) para generar reacciones en su audiencia virtual. Por ese motivo, este proyecto es pionero en esta parte, ya que aplica análisis de las emociones transmitidas por el orador para posteriormente generar reacciones en tiempo real en el público virtual, de forma que el orador reciba retroalimentación y pueda modificar su discurso de forma acorde.

Hacer análisis de emociones es un proceso complejo y costoso. Por lo tanto, entre las metas de este proyecto no se incluye la implementación de un sistema de análisis de emociones, pero sí se hace uso de varios ya implementados, centrando el foco en dos cuestiones principales: el análisis de la voz del orador, sin tener en cuenta el contenido, y el análisis del contenido del mensaje transmitido.

Antes de proceder al reconocimiento automático de emociones, es necesario decidir cómo se van a representar las emociones que se detecten. Hay dos métodos básicos [9] para definir emociones: dimensiones emocionales y categorías emocionales.

Por un lado, las **dimensiones emocionales** se encargan de representar los aspectos importantes de los conceptos emocionales, es decir, se clasifican las emociones según si son positivas o negativas, activas o pasivas y el control que ejerce la emoción.

Por otro lado, las **categorías emocionales** son la forma de definir las las emociones mediante palabras o etiquetas (e.g. alegría, entusiasmo, agresividad, tristeza, etc.). Dado que existen muchísimas palabras para describir los estados emocionales, en muchos sistemas se reducen las emociones a seis grupos [11] de **emociones básicas**: alegría, tristeza, sorpresa, miedo, enfado y emoción neutra.

1.1.2.1. Análisis de voz sin tener en cuenta el contenido

Según Laukka [15], tanto cuando se habla como cuando se imparte una conferencia, **el mensaje que se transmite va acompañado de las emociones del individuo**. Dichas emociones afectan al aparato respiratorio de forma que el tono de la voz se modifica. A partir de estos cambios en el tono de voz y sin tener en cuenta el contenido, se puede establecer cuál es la emoción que está afectando a este discurso. Evidentemente, no todas las emociones que siente el orador se transmiten o reflejan en el tono de voz pero, a efectos de este trabajo, lo que interesa son aquellas emociones que sí puede percibir el público y, en consecuencia, reaccionar a ellas.

Hay muchas investigaciones recientes que se han centrado en el recono-

cimiento automático de emociones en la voz. La mayoría de estos trabajos [34] se basan en el uso de redes neuronales, máquinas de soporte vectorial o modelos gaussianos que se entrenan con grandes volúmenes de grabaciones de audio previamente clasificadas. Estas grabaciones sirven para entrenar los sistemas de predicción y, posteriormente, se hacen pruebas enviando audios sin clasificar para ver la cantidad de aciertos que obtiene el sistema. Se hacen sobre esto varias iteraciones, evaluando cada una de ellas, hasta que el modelo haga predicciones con una tasa mínima de error y sin llegar al sobreajuste.

Existen muchos volúmenes [54, 12, 57] de datos con grabaciones de audio emocionales. La creación de estos volúmenes de datos, con los que se entrenan las redes neuronales, se puede hacer grabando a alguien hablando de forma espontánea, grabando conversaciones entre individuos (actuación no motivada) o a actores que interpretan frases con una emoción determinada (actuación motivada).

La teoría [47] dice que, dependiendo del estado emocional, del orador el tono de voz cambia. Por ejemplo, cuando un individuo está triste, tiende a tener un tono más grave y un tempo más lento. Sin embargo, cuando alguien está feliz, el tempo se acelera y el tono de voz se agudiza.

1.1.2.2. Análisis del contenido del mensaje transmitido

La emoción que genera la información que el orador quiere transmitir influye decisivamente en la selección de las palabras y la estructura de las frases que expresa.

Actualmente existen diferentes metodologías para detectar la emoción a través del texto, las cuales se pueden dividir en cuatro apartados [15]: detección de palabras clave, afinidad léxica, procesamiento estadístico del lenguaje natural y métodos basados en el conocimiento del mundo real.

En la **detección de palabras clave** en el texto, las emociones se extraen basándose en la presencia de palabras que hacen referencia a emociones o palabras afectivas.

La **afinidad léxica**, además de detectar palabras afectivas evidentes, asigna a otras palabras la probabilidad de ser dichas cuando el sujeto experimenta una determinada emoción.

El **procesamiento estadístico** consiste en utilizar una red neuronal entrenada sobre una base de datos muy grande de textos clasificados con las emociones que representan. Este último método tiene en cuenta cualquier

tipo de palabra e incluye también signos de puntuación para detectar las emociones.

Los métodos basados en el **conocimiento del mundo real**, además de analizar los aspectos del texto, evalúan las características afectivas del contenido semántico subyacente del texto. Estas técnicas permiten obtener las emociones del contenido semántico a pesar de que no existan palabras que hagan referencia a emociones. Para ello se basan en conocimientos del mundo real, como las actitudes de los individuos en ciertas situaciones.

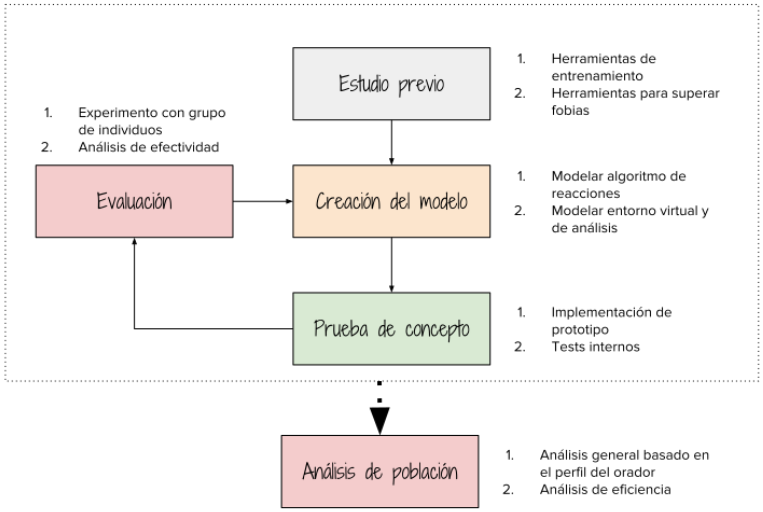

Figura 1.2: Diagrama de objetivos

1.2. Objetivos

El objetivo primordial de este proyecto es construir un videojuego de realidad virtual con dos propósitos.

> Por un lado, proporcionar una **herramienta reactiva para oradores**, de forma que éstos puedan practicar sus presentaciones ante una audiencia virtual que reacciona en tiempo real a las características del orador y le proporciona retroalimentación. Así el orador puede ensayar sus presentaciones y evaluar su discurso, para mejorarlo o habituarse a modificar la ponencia según las reacciones de la audiencia.
>
> Por otro lado, servir de apoyo a terapias para el **tratamiento de la patología de la glosofobia**, proporcionando un entorno seguro en el que empezar a enfrentarse al problema de la evitación, con el fin de disminuir el miedo a hablar en público, como paso previo a la exposición real necesaria para superar este miedo.

Para ello se creará una experiencia en un entorno de realidad virtual controlado para que un orador pueda hablar frente a una audiencia virtual.

Para conseguir esto se ha dividido el trabajo en cuatro subobjetivos, como se puede apreciar en la figura 1.2: hacer un estudio de los trabajos previos, crear el prototipo de los entornos virtual y de análisis, generar reacciones en la audiencia en tiempo real y realizar varios experimentos que ayuden a implementar el proyecto y a validar la efectividad de este trabajo.

1. **Hacer un estudio de los trabajos previos**

   Antes de proceder con la implementación del entorno, es necesario investigar y conocer qué otras herramientas hay en el mercado que traten el mismo tema o cuestiones similares, además de ver qué aplicaciones tiene la realidad virtual y en qué contextos se ha usado.

2. **Crear los entornos virtual y de análisis**

   En este entorno se posiciona al usuario en situaciones diversas lo más realistas posibles (e.g. un auditorio grande para dar una conferencia, una clase o una entrevista de trabajo), permitiéndole entrenar su discurso y desarrollar sus habilidades como orador para poder afrontar mejor este tipo de situaciones en la vida real. La experiencia se creará a modo de videojuego para mejorar la inmersión, la persecución de objetivos y el aprendizaje. Con esto se pretende conseguir que el usuario tenga la sensación de estar en un entorno de juego y disminuir la sensación de miedo escénico que puede sufrir en la misma situación en la vida real. La experiencia incluirá diferentes elementos de jugabilidad para fomentar la motivación del orador por seguir con la experiencia: sistema de puntuación, porcentaje de efectividad del discurso, cantidad de público atento a la presentación, etc. De esta manera, el orador tendrá en todo momento pistas visuales y auditivas sobre el estado de la audiencia, que le permitirán adaptar su discurso en tiempo real. Además de eso, al finalizar el juego, recibirá un informe más detallado con estadísticas de cómo ha ido su presentación y los porcentajes de efectividad segmentados por tiempo. En este punto existe la posibilidad de permitir al orador visualizar la grabación de su charla para ver qué partes de su entrenamiento han sido más o menos efectivas, lo que le servirá de aprendizaje y refuerzo.

3. **Generar reacciones en la audiencia en tiempo real**

   Saber cuál es la reacción de un público ante una presentación, o determinadas partes de ella, ofrece al orador la oportunidad para adaptarse y reaccionar en tiempo real, igual que si lo hiciera delante de un público real, pero con la salvedad de que lo hace en un entorno seguro. Esto le permite prepararse con antelación a esas reacciones o modificar su discurso para que la audiencia mejore su reacción.

Por lo tanto, uno de los pilares básicos de este proyecto es su audiencia virtual reactiva. Lo que se pretende, mediante el análisis de diferentes parámetros del orador (tono y proyección de la voz, postura o hacia dónde dirige la mirada), es provocar reacciones en la audiencia virtual de forma que el orador tenga realimentación en tiempo real sobre la efectividad de su discurso. Este tipo de realimentación, más que otras posibles (puntuaciones, alarmas de juego,...), es mucho más favorable a la inmersión, ya que simula lo que ocurre en la realidad cuando damos una charla. Las reacciones se deben generar en base a las emociones que se extraen de los parámetros o características del orador, junto con un modelado del público para tratar de predecir cuál sería el impacto generado en un público real a partir de las acciones del orador.

4. **Realizar experimentos**

   Para saber qué tipo de emociones se generan en una audiencia real en función de las acciones del orador, uno de los objetivos es realizar un experimento para analizar cómo afectan determinadas características de varios oradores a una audiencia real compuesta por un grupo de individuos, para así poder aplicarlo a las reacciones de la audiencia virtual. Otra meta importante es realizar un experimento de un primer prototipo con un grupo diverso de individuos, para poder sacar conclusiones sobre su funcionamiento y poder afinar las reacciones de la audiencia virtual. En este experimento será importante intentar conseguir la sensación de presencia en los oradores, para así demostrar que los oradores experimentan un miedo escénico similar frente a una audiencia virtual que frente a una real.

1.3. Estructura del documento

Este documento está organizado en 6 capítulos, cada uno dedicado a una temática relacionada con este trabajo. Esta sección está encuadrada en el presente capítulo 1 de Introducción. El capítulo 2 recopila todo el trabajo de estudio previo realizado antes de abordar la implementación. El capítulo está dividido en dos subsecciones: un apartado dedicado a herramientas de entrenamiento y simuladores (2.1) y otro apartado dedicado a herramientas para tratar con fobias o miedos (2.2).

El capítulo 3 de Creación del entorno engloba la parte más importante de este documento, en la que se hace una descripción detallada del sistema (3.1) y se explican todos los componentes del sistema y su arquitectura (3.2). Posteriormente, en la sección 3.3 se describen las tecnologías investigadas para la implementación y cuáles de ellas han sido elegidas para el desarrollo de este trabajo.

En el capítulo 4 de Experimentos se puede visualizar una descripción del experimento realizado en el contexto de este trabajo. En este capítulo se describe el objetivo del experimento (4.1), la metodología usada (4.2), los resultados obtenidos (4.3) y una discusión sobre estos resultados (4.4).

Este trabajo tiene una peculiaridad, pues cuenta con un capítulo específico (5) e inusualmente grande de trabajo futuro. Esto es debido a que se pretenden reflejar en este capítulo muchas de las ideas que han surgido durante este Trabajo de Fin de Máster, las cuales se pretenden continuar en el marco del proyecto ComunicArte (5.1), un proyecto más ambicioso, financiado por parte de la Fundación BBVA.

Por último, se exponen las conclusiones a las que se ha llegado después de realizar este trabajo en el capítulo 6 de Conclusiones.

Chapter 1

Introduction

Since ancient times, the need of public speaking has been a constant for human kind. Public speaking has been used throughout history to persuade, convince, teach or event to conduct the thinking of others.

The Greeks were pioneers in this matter, first the sophists like Protagoras [48], and after that great orators like Demosthenes and Socrates [60]. They originated the rhetoric, a set of disciplines and a specific way of structuring a speech with a final goal: persuading the audience. The **rhetoric** is transversal to different knowledge fields (literature, politics, journalism, education, social sciences, law, biblical studies, etc.) and takes care of studying and standardizing language techniques and procedures with an esthetical or persuasive purpose, apart from its communicative one.

Afterwards, the Romans (i.e. Cicero [7] and Quintilian [44]) followed their predecessors and continued perfecting this discipline with their publications about it.

Approaching our time, many personalities have stood out for their use of the oratory to lead the the course of events: among them was Abraham Lincoln, with his Gettysburg speech [8], Martin Luther King and his *I Have a dream* [27], Winston Churchill, Nelson Mandela or Steve Jobs.

Oratory and the art of public speaking with eloquence, in order to persuade or affect the audience, is a transversal ability used in many areas. This skill has been important for the human beings throughout history. Oral communication also includes what is not said (body language, rhythm, tone of voice, etc.).

It is, therefore, a cross-disciplinary practice to different areas of human

life: apart from giving a conference before an audience, it can also be put into practice to speak at a neighbors' meeting, intervene in class, give a point of view and defend it, speak in front of an HR manager at a job interview or even give a speech at a wedding.

However, the cultivation of this discipline has faced an obstacle that has also accompanied human kind throughout history: the fear of public speaking.

According to the Spanish Society for the Study of Anxiety and Stress (SEAS) [39], when speaking in public, most people react with high levels of activation, due to the effort involved in the cognitive management of information (recovery of the information, relating some elements with others, etc.), the verbalization of the contents, with an increased volume of voice, maintaining a high expenditure of resources (energy, attention, etc), for a long time. This set of factors is what is experienced as nerves or stage fright. **Stage fright** is not new to those familiar with the performing arts. However, in the theater, for example, actors can take refuge behind the character they embody to mitigate its effects. But these techniques are not appropriate to give talks, since one of the premises of the good speaker is to "be yourself".

A significant percentage of people are activated too much and can't avoid focusing almost all their attention on their symptoms of anxiety and not in their performance. This makes the situation result in a tremendously unpleasant experience, thus, they try to avoid it at all costs. Some people who practice avoidance end up receiving a diagnosis of anxiety disorder due to **social phobia**.

According to the Royal Spanish Academy (RAE), fear is:

1. m. Anguish over a real or imagined risk or damage.
2. m. Distrust or apprehension that someone has something that happens to him contrary to what he wants.

and scenic refers to:

1. adj. Pertaining to or relating to the scene.

It can be assumed, then, that stage fright is an anguish because an event (real or imaginary) happens when one is on a stage. There are many ways to call this fear: stage fright, fear of public speaking, etc. Sometimes, this stage fright leads to a phobia. Again, according to the RAE, phobia is:

1. f. Exaggerated aversion to someone or something.
2. f. Psychiatrist. Anguished and uncontrollable fear of certain acts, ideas, objects or situations, which is absurd and approaches obsession.

Scenic fear contemplates other public performances in addition to talking such as singing, acting, dancing, etc. However, this document focuses only on the fear of public speaking, this one being more specific than stage fright.

The technical name for the phobia of public speaking is **glossophobia** (anxiety to speak). This alteration differs from social phobia in the sense that in the latter the person fears any type of activity that requires socialization, while in glossophobia the feared element is only the activity of public speaking [17]. The fear that a person with glossophobia feels comes up whenever that person faces the activity of speaking in front of an audience and is independent of the context and the particular situation. For this type of phobia it is not enough to apply exposure techniques, but also professional help is necessary.

The fear of public speaking, which is born of shyness or simply of the anxiety caused by the situation of facing an audience, is very widespread among the population. There is a test called the Trier test, created in the nineties to measure social stress.

> This test exploits precisely the fear of public speaking and consists of a series of volunteers being taken to a room where three people tell them they have ten minutes to prepare a five-minute presentation. There is a camera that will record the presentation and after that the subjects must count from number 1022 to 13 backwards, loudly and without committing failures.

This type of experiment puts the subject in a situation of anticipation of danger, so that a reaction, generally intense, arises as a result of catastrophic anticipatory thoughts about the real or imagined situation of public speaking.

The fear of public speaking can be defined as a disproportionate response of the body to a situation, confusing worry and nerves with a threat. In this way, the heart rate accelerates and may be accompanied by other symptoms such as sweating, tremors, shortness of breath, reddening of the skin, loss of the narrative thread and even dizziness. **The body anticipates in this way a threatening situation that is not real**.

Therefore, as public speaking is a necessary skill in the day to day of people and knowing that the fear of public speaking is a condition suffered

by around 75% [52] of the population, it is natural to think of looking for some way to alleviate the problem [50].

1.1. Motivation and objectives of the work

At present, the need to speak in front of an audience occurs in different areas of human life, from making a presentation at an event to speaking in a job interview. In fact, in the Horizon 2020[1], a Framework Program that integrates part of the activities of the European Union of research and innovation in the period 2014-2020, teaching Europeans public speaking is one of the sections that has had great importance.

Public speaking eloquently and convincing, moving or persuading the audience is a cross-cutting skill that must be used in various situations, such as those mentioned above.

However, stage fright and fear of public speaking are among the most prevalent fears in the world and affect a large part of the population. They are, therefore, a major obstacle to developing this skill.

As can be seen in the next chapter, all applications with the same purpose as this work, face the challenge with greater or lesser success, but in none of them exist the two elements that in this work are considered crucial: **offering feedback to the speaker and making him improve his presentation skills while he is doing it**.

In the first place, as we will see, only two of the researched tools incorporate sensors to measure some of the characteristics of the speaker: his speech and gestures, his volume and tone of voice, where does he look or his posture.

Secondly, and more importantly, none of the tools in the literature reviewed offer a reactive audience in real time, based on the data obtained from these sensors.

For this reason, the idea of this work is to combine virtual reality and sensors that measure the actions of the speaker, allowing the system to react and offer feedback in real time, so that the speaker can modify his speech accordingly. In addition, the user is offered a **gamification** of the environment, a game mode, and not so much a therapy or simulator to use, so that the experience is immersive and encourages motivation to improve.

[1] `https://eshorizonte2020.es/`

1.1.1. Virtual reality and VRET

Virtual reality is a realistic three-dimensional environment, composed of images or 3D models, which is created through the use of interactive hardware and software and which is presented to the user so that a user can accept it as a real environment, in which he interacts in a similar or equal way to the physical world.

In recent years, virtual reality has experienced a great development thanks to the growing availability of virtual reality devices.

There are several studies [50, 41] that support the use of virtual reality in the context of the treatment of phobias and fears that individuals experience. The premise that must be met in this type of treatments is the so-called **presence** response [49, 3, 16], which is a way of measuring the capacity of an interactive experience in virtual reality to make individuals feel "really" in a virtual environment, meaning that the degree of anxiety experienced by the subject within the virtual reality environment must be sufficiently similar to how he would react normally in a similar situation in the real world.

The most common forms of presence are telepresence, which is the capacity of a virtual environment to create an illusion of "being there", and social presence, which defines the feeling of "being there with another" and analyzing, for example, the interactions between the real and the virtual.

For this project it is interesting to explore both aspects of the presence, because public speaking is a social activity by definition.

Immersion is the perception of being physically within a non-physical world. It encompasses the sense of presence, which is the point at which the human brain believes it is in a place where it is not. This is achieved through purely mental and / or physical means. The state of total immersion exists when there are enough senses activated to create the perception of being present in a virtual world. There are two common types of immersion:

1. Mental immersion: A deep mental state in which the belief that one is in a virtual environment is suspended.
2. Physical immersion: State in which physical interaction with the virtual environment occurs.

The objective of virtual reality is to generate total immersion, that is, to generate a sensory experience that seems so real that the user forgets that he is inside a virtual environment created artificially, and interacts with it

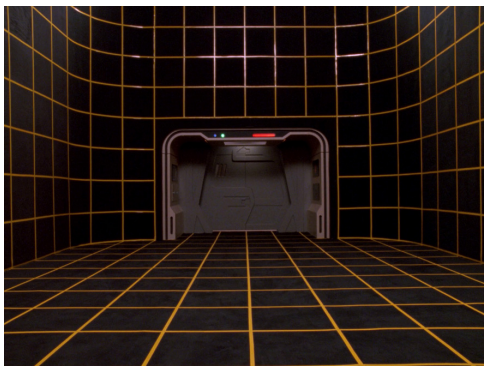

Figure 1.1: Star Trek: Holodeck

as he would in the real world. In a virtual reality environment the world is completely synthetic and may or may not mimic the properties of a real environment. This implies that virtual reality can simulate a daily situation (i.e. walking through the streets of a city, interacting with an office environment, etc.), or it can exceed the limits of the physical creating a world in which the laws of Physics such as gravity, time and properties of materials do not hold up (i.e. shooting aliens on a planet without gravity).

Virtual reality requires that as many of our senses as possible to be stimulated. The correct stimulation of these senses requires sensory feedback, which is achieved through integrated hardware and software. Examples of this hardware can be the HMD (*head-mounted display*), special gloves with tactile feedback or hand or control accessories [46].

For the human brain to accept an artificial and virtual environment as real, not only does this virtual world need to appear real, but it also has to "feel" real. Making the world look real can be achieved through the HMD, which recreates a life-like 3D virtual environment without the limitations of a screen. To make the world feel real, system inputs can be used such as movement followers that base interactivity on user movements. If all the senses that are normally used to function in the real world are stimulated, virtual reality environments begin to give the same sensation as the real world.

One of the main applications of virtual reality is in the world of videogames, due to the immersive capacity of virtual reality mentioned above. This application is very widespread in the different types of existing video games, from first-person adventure games to graphical adventures with fiction narrative.

But this concept is not new at all. In fact, already in 1997 Janet Murray

[36] ran as an advocate of the idea that computers could one day offer interactive experiences so realistic that they would become indistinguishable from reality. This author imagined a future in which a computer could offer perfect narratives within the Holodeck[2], a fictional device consisting of a room in which Star Trek characters entered to relax. In this room, the computer created a virtual reality environment in which the participants became the protagonists of a novel. Depending on the interactions they had with the characters in the story, the computer was creating the continuation of the novel. For Murray, becoming a character in a fiction within the Holodeck would be a pleasant experience as much as learning. As in a video game, the Holodeck offers a safe place to face annoying feelings that are otherwise suppressed; it allows a user to recognize the most terrifying fantasies without being paralyzed by them [25].

However, video games are not the only application of virtual reality. There are many areas in which this technology is used. For example, in the military field, virtual reality is used for flight simulators or war simulators, and in the health area, virtual reality is used to simulate medical operations environments. There are many more cases, but these two examples demonstrate the ability of virtual reality to **generate safe environments where to practice certain skills**, be it flying an airplane or a tank, doing a heart transplant, reacting in case of flight in a nuclear power station or even, as this work suggests, practice the ability of public speaking. Basically, the concept that is being handled here is to create a safe environment in which a user can fail without this having consequences in the real world. A fundamental concept for the learning experience to occur.

Regardless of their type, virtual reality experiences must be designed differently than usual videogames, since the movements of the head in the virtual world can produce a sickness effect known as simulator disease or **simulator sickness**. This effect was documented for the first time in a helicopter simulator by Havron and Butler in 1957 [19] and refers to the discomfort generated by the body when it experiences a virtual movement that has no place in the real world.

VRET (*Virtual Reality Exposure Therapy*) is using virtual reality in occupational or psychological therapy [53]. In this type of therapies, patients navigate a world of virtual reality, created digitally and designed specifically to perform tasks that allow to treat specific pathologies.

In VRET, an individual is introduced into a virtual environment generated by computer, either using an HMD or entering a room with screens around. This environment can be programmed so that the person can face

[2] `http://www.startrek.com/database_article/holodeck`

situations or places that he or she fears and that may not be safe to confront in the real world.

There are some studies [59, 45] that show that VRET can be useful to treat different anxiety disorders or problems related to anxiety, including claustrophobia, fear of driving, fear of heights, fear of flying, fear of spiders or social phobia. There are also studies [45] that use VRET for the treatment of post-traumatic stress disorders.

1.1.2. Emotion analysis

In addition to the virtual environment, it is important to perform an analysis of the speaker's emotions, to determine if he or she is experiencing that sense of presence and to gauge the reactions of the virtual audience.

Although in the VRET area there is a lot of research regarding its use to deal with the fear of public speaking, in the area of ??modeling a virtual audience it has not been possible to find any work that does an analysis of the speaker's parameters (where does he look, his tone and projection of voice, his gestures, etc.) to generate reactions in the virtual audience. For this reason, this project is a pioneer in this part, since it applies analysis of the emotions transmitted by the speaker to later generate reactions in real time in the virtual audience. Through this reactions the speaker receives feedback and can modify his speech accordingly.

Doing emotions analysis is a complex and expensive process. Therefore, the goals of this project do not include the implementation of an emotional analysis system, but several already implemented tools are used, focusing on two main issues: the analysis of the speaker's voice, without taking into account the content of the speech, and the analysis of the content of the message transmitted.

Before proceeding to the automatic recognition of emotions, it is necessary to decide how the emotions that are detected will be represented. There are two basic methods [9] of defining emotions: emotional dimensions and emotional categories.

On the one hand, **the emotional dimensions** are responsible for representing the important aspects of emotional concepts, that is, the emotions are classified according to whether they are positive or negative, active or passive, and the control exerted by the emotion.

On the other hand, **emotional categories** are the way to define emo-

tions through words or labels (e.g. joy, enthusiasm, aggression, sadness, etc.). Since there are many words to describe emotional states, in many systems emotions are reduced to six [11] groups of **basic emotions**: joy, sadness, surprise, fear, anger and neutral emotion.

1.1.2.1. Analysing voice without taking into account the speech content

According to Laukka et al. [15], when public speaking and when a lecture is given, **the message that is transmitted is accompanied by the emotions of the individual**. These emotions affect the respiratory system and the tone of the voice changes. From these changes in the voice tone and without taking into account the content, it can established what emotions are affecting this speech. Obviously, not all the emotions that the speaker feels are transmitted or reflected in the tone of voice, but, for the purposes of this work, the main focus are those emotions that the public can perceive and, consequently, react to them.

There are many recent investigations that have focused on the automatic recognition of emotions in the voice. Most of these works [34] are based on the use of neural networks, vector support machines or Gaussian models that are trained with large volumes of previously classified audio recordings. These recordings are used to train the prediction systems and, later, tests are done sending unclassified audios to see how many hits the system gets. Several iterations are made on this, evaluating each of them, until the model makes predictions with a minimum error rate and without reaching the overfit.

There are many volumes [54, 12, 57] of data with emotional audio recordings. The creation of these data volumes, with which neural networks are trained, can be done by recording someone speaking spontaneously, recording conversations between individuals (unmotivated performance) or actors who interpret sentences with a certain emotion (motivated acting) .

The theory [47] says that, depending on the emotional state, the tone of voice changes. For example, when an individual is sad, he tends to have a lower tone and a slower tempo. However, when someone is happy, the tempo accelerates and the tone of voice sharpens.

1.1.2.2. Analysing the content of the transmitted message

The emotion generated by the information that the speaker wants to convey has a decisive influence on the selection of the words and the structure

of the sentences he expresses.

Currently there are different methodologies to detect emotion through text, which can be divided into four sections [15]: keyword detection, lexical affinity, statistical processing of natural language and methods based on knowledge of the real world.

In the **detection of keywords** in text, emotions are extracted based on the presence of words that refer to emotions or affective words.

The **lexical affinity**, besides detecting evident affective words, assigns to other words the probability of being said when the subject experiences a certain emotion.

Statistical processing consists of using a neural network trained on a very large database of texts classified with the emotions they represent. This last method takes into account any type of word and also includes punctuation signs to detect emotions.

The methods based on the **knowledge of the real world**, in addition to analyzing the aspects of the text, evaluate the affective characteristics of the underlying semantic content of the text. These techniques allow to obtain the emotions of the semantic content even if there are no words that refer to emotions. For this part, they base their analysis on knowledge of the real world, such as the attitudes of individuals in certain situations.

1.2. Objectives

The main objective of this project is to build a virtual reality video game with two purposes.

> On the one hand, provide a **reactive tool for speakers**, so that they can practice their presentations before a virtual audience that reacts in real time to the features of the speaker's speech and provides feedback. Thus, the speaker can rehearse his presentations and evaluate his speech, to improve it or get used to modifying the presentation according to the reactions of the audience.
>
> On the other hand, support therapies for the **treatment of the pathology of glossophobia**, providing a safe environment in which to begin to face the problem of avoidance, in order to reduce the fear of speaking in public, as a previous step to the real exposure necessary to overcome this fear.

This work pretends to create an experience in a controlled virtual reality environment for a speaker to speak in front of a virtual audience.

To achieve this, the work has been divided into four sub-objectives: doing a study of the previous works, creating the prototype of the virtual and analytical environments, generating reactions in the audience in real time and carrying out several experiments that help to implement the project and validate the effectiveness of this work.

1. **Doing study of the previous works**

   Before proceeding with the implementation of the environment, it is necessary to investigate and know what other tools exist in the market that deal with the same issue or similar issues, as well as seeing what are virtual reality applications and in what contexts it has been used.

2. **Creating the virtual and analytical environments**

   In this environment the user is positioned in diverse situations as realistic as possible (i.e. a large auditorium to give a lecture, a class or a job interview), allowing him to train his speech and develop his skills as a speaker to better cope with this type of situations in real life. The experience will be created as a video game to improve immersion, the pursuit of objectives and learning. This is intended to ensure that the user has the feeling of being in a game environment and reduce

the sense of stage fright that can suffer in the same situation in real life. The experience will include different elements of gameplay to encourage the motivation of the speaker to continue with the experience: scoring system, percentage of effectiveness of the speech, number of public attentive to the presentation, etc. In this way, the speaker will have at all times visual and auditory clues about the state of the audience, which will allow him to adapt his speech in real time. In addition to that, at the end of the game, he will receive a more detailed report with statistics of how his presentation has progressed and the percentages of speech effectiveness segmented by time. At this point there is the possibility of allowing the speaker to view the recording of his talk to see what parts of his training have been more or less effective, which will serve as learning and reinforcement.

3. **Generating reactions in the audience in real time**

   Knowing what the reaction of an audience to a presentation, or certain parts of it, offers the speaker the opportunity to adapt and react in real time, just as if he would do in front of a real audience, but with the exception that he does it in a secure environment. This allows you the speaker to prepare in advance for those reactions or modify his speech so that the audience improves their reaction.

   Therefore, one of the basic pillars of this project is its reactive virtual audience. What is intended, through the analysis of different parameters of the speaker (tone and projection of the voice, posture or where does he look), is to provoke reactions in the virtual audience so that the speaker has feedback in real time on the effectiveness of his speech. This type of feedback, more than other possible types (scores, game alarms, ...), is much more favorable to immersion, since it simulates what happens in reality when we give a talk. The reactions must be generated based on the emotions that are extracted from the parameters or characteristics of the speaker, together with a modeling of the audience to try to predict what the impact would be on a real audience based on the speaker's actions.

4. **Carrying out experiments**

   To know what kind of emotions are generated in a real audience based on the actions of the speaker, one of the objectives is to conduct an experiment to analyze how certain features of several speakers affect a real audience composed of a group of individuals, in order to be able to apply it to the reactions of the virtual audience. Another important goal is to conduct an experiment of a first prototype with a diverse group of individuals, in order to draw conclusions about its operation and to refine the reactions of the virtual audience. In this experiment it will be important to try to get the feeling of presence in the speakers,

in order to demonstrate that the speakers experience a similar stage fright in front of a virtual audience versus a real audience.

Capítulo 2

Estado del arte

En esta sección se hace un repaso de algunos de los trabajos que tratan la misma temática o persiguen los mismos objetivos que el proyecto Hola UniVRso. Para ello, en el apartado 2.1, se analizan primero proyectos con simuladores o herramientas de entrenamiento diseñadas para permitir al usuario aprender y entrenar diversas habilidades. Posteriormente, en el apartado 2.2, se analizan algunas herramientas y metodologías para tratar o paliar diferentes fobias y miedos, tanto por medio de acciones que no involucren la tecnología (sección 2.2.1) como por otras que sí lo hagan (sección 2.2.2).

2.1. Herramientas de entrenamiento y simuladores

Existen multitud de herramientas en el mercado planteadas para el aprendizaje y perfeccionamiento de diversas habilidades. Aunque las aplicaciones pioneras en entrenamiento usando la realidad virtual son los simuladores, existen multitud de otros tipos de aplicaciones dedicadas al entrenamiento y aprendizaje. A continuación se ejemplifican algunos de los ámbitos más importantes en los que se usan aplicaciones de este tipo.

El primer simulador de vuelo fue patentado en 1930 por la compañía Link[1] en Canadá. Estaba basado en los mecanismos neumáticos de pianos y órganos. Se fabricaron más de 7000 dispositivos para entrenar a los pilotos de la Segunda Guerra Mundial. A medida que la tecnología evolucionaba, los simuladores iban mejorando. Con la aparición de la realidad virtual, las posibilidades de aplicación de los simuladores se han incrementado de forma notable.

En el ámbito militar, la realidad virtual se lleva usando desde hace décadas. Desde los años 50, Estados Unidos ya cuenta con sistemas de entrenamiento militar [37] usando la realidad virtual, donde los soldados pueden enfrentarse a diversas situaciones de combate en entornos seguros o practicar habilidades de pilotaje de aviones, tanques o buques de la marina. La realidad virtual se ha convertido en un medio de entrenamiento para personal militar, en el que se puede aprender a reaccionar de la mejor forma sin el riesgo de las situaciones reales. Además, este método se ha demostrado más seguro y menos costoso que los métodos tradicionales.

- **War Thunder**[2]

  War Thunder es un simulador vuelo de combate que tiene lugar en la Segunda Guerra Mundial. Permite al jugador elegir el tipo de aeronave o vehículo a pilotar y consiste en pilotar dicho vehículo y reaccionar correctamente a las diferentes situaciones que se van presentando para no ser derribado. El juego es multijugador y permite competir contra otros jugadores conectados al mismo tiempo.

- **X-Plane 11**[3]

  Esta herramienta es un simulador de vuelos de aviones comerciales que incluye multitud de modelos de aviones y réplicas 3D de más de 3000 aeropuertos del mundo. Mediante este simulador, se puede hacer el

[1] `http://link.com/`
[2] `https://warthunder.com/es`
[3] `https://www.x-plane.com/`

entrenamiento de pilotaje, despegue y aterrizaje, manejo de controles del avión y cualquier otra cuestión relacionada.

En el área de la salud, por las mismas ventajas que ofrece en el ámbito militar, también se está usando cada vez más la realidad virtual para simular operaciones quirúrgicas [42] de riesgo o aprender a realizar tareas cotidianas de tratamiento de pacientes y diagnóstico [14].

- **VirSSPA**[4]

  Esta herramienta es una aplicación médica que usa la realidad virtual para la optimización y planificación de la cirugía. Permite la generación de un modelo del paciente en un entorno virtual para ser intervenido virtualmente a partir de imágenes radiológicas. La aplicación dispone de las herramientas que se usan normalmente en quirófano, posibilitando la simulación de la operación.

También se hace uso de la realidad virtual para entrenar procesos de fabricación en la industria, de forma que se aprenden, por ejemplo, mecánicas de montaje complejas sin las implicaciones de hacerlo en un entorno real. También, en este mismo ámbito, se puede entrenar a hacer reparaciones o seguridad y gestión de riesgos.

- **ITI Crane Simulator**[5]

  Se trata de un simulador de manejo de grúas que se usa para enseñar a los aprendices que van a usar este vehículo a manejar la grúa y controlar la seguridad en el transporte de materiales.

- **Unigine Space**[6]

  Simulador de varias misiones espaciales que permite el entrenamiento en la reparación y manejo de controles de satélites, vehículos espaciales y cohetes. Oferece una virtualización de un entorno en el espacio, dibujado de trayectorias y representación de cuerpos celestes.

Por último, en la educación se está usando la realidad virtual como herramienta complementaria para fomentar el aprendizaje [43], aumentar la motivación de los alumnos, para que estos puedan retener conceptos más

[4] `http://www.itemas.org/cartera-de-servicios/transferencia-de-tecnologia/itt/virsspa/c/show/`

[5] `https://www.iti.com/vr/carry-deck-training-simulator`

[6] `https://unigine.com/en/industries/simulation/space`

rápido, o experimentar de forma práctica situaciones que, enseñadas de forma teórica resultan tediosas. Por ello, existen proyectos que usan la realidad virtual en el ámbito educativo para tareas muy diversas, ya sea una simulación de épocas pasadas para aprender historia, un sistema de formación de tormentas y fenómenos meteorológicos o visitar de forma virtual un museo.

- **Unimersiv**[7]

  Unimersiv es una herramienta educativa en la que es posible explorar usando realidad virtual diferentes escenarios para aprender sobre ellos. Por ejemplo, disponen de una experiencia para aprender sobre dinosaurios, un viaje virtual al cerebro humano o una experiencia sobre la construcción del Acrópolis de Atenas.

- **ClassVR**[8]

  Una herramienta que pone realidad virtual al servicio de las clases, para simular entornos, hacer lecciones prácticas, y en definitiva, aportar realismo a algunas lecciones para favorecer el aprendizaje.

Todas las aplicaciones anteriores son meros ejemplos del gran abanico de aplicaciones y herramientas existentes para entrenamiento y aprendizaje usando la realidad virtual. La gran disponibilidad de herramientas que usan la realidad virtual para entrenar habilidades pone de manifiesto que es un sistema efectivo para ello, ya que se trata de una técnica que proporciona un entorno más seguro y que permite representar cualquier situación imaginable.

[7] `https://unimersiv.com/`
[8] `http://www.classvr.com`

2.2. Herramientas para tratar con fobias o miedos

2.2.1. Sin utilizar la tecnología

Cuando una persona se enfrenta a la tesitura de tener que hablar en público y es consciente de su miedo, puede recurrir a técnicas más allá de la simple preparación de unas diapositivas y entrenamientos al uso. Existen diversas formas y técnicas para lidiar con el miedo a hablar en público.

Por un lado están las terapias [56] con profesionales, en las que uno acude a un psicólogo para que, a través de varias sesiones, éste le ayude a superar el miedo a hablar en público. Algunas de estas terapias se centran en la **reestructuración cognitiva**, es decir, mostrarle al paciente que el estrés y la ansiedad se producen en su interior, de forma que la mente engrandece la amenaza y se centra exclusivamente en pensar en el problema que supone dicha amenaza.

Otras técnicas se basan en la **relajación**, con la intención de lograr reducir esa aceleración de ritmo que se produce en el cuerpo cuando uno se encuentra en la situación de exponer ante un público.

Otro tipo de **abordaje es de carácter más grupal**, como los talleres para hablar en público, en los que se entrenan habilidades sociales, se dan trucos que funcionan a la hora de dar una charla y se practica delante de público reducido para ir perdiendo el miedo.

2.2.2. Utilizando la tecnología

El hecho de utilizar la tecnología para el tratamiento de fobias es una idea que se ha llevado a la práctica en multitud de ocasiones.

Esto se debe a que la tecnología permite simular entornos seguros en los cuales la persona con una determinada fobia puede enfrentarse a ella, sabiendo que la simulación está bajo su absoluto control y que en cualquier momento la puede parar.

La realidad virtual, aunque en fase experimental [38], se usa ya para ayudar en otros entornos similares como la recuperación neurológica o los tratamientos de trastornos por estrés traumático. Evidentemente, es una herramienta complementaria al uso de terapia psicológica cognitivo-conductual, puesto que aunque puede ayudar, no sustituye la necesidad de exposición real en algunos casos. Sin embargo, en el campo de VRET, a pesar de que hay

estudios sobre el tema de hace 20 años, todavía no hay un uso claro de esta técnica para casos reales.

Más concretamente, para tratar el miedo escénico o miedo a hablar en público, existen en el mercado algunas aplicaciones que tienen acercamientos más o menos similares a lo que se plantea en este documento.

A continuación, se hace una enumeración de las aplicaciones anteriormente mencionadas, destacando sus puntos fuertes y su forma de afrontar el reto. Casi todas estas herramientas han sido probadas a lo largo de una jornada íntegramente dedicada a ello por parte de un grupo de aproximadamente 15 personas. En esta jornada, además de probar las aplicaciones, se han intercambiado impresiones sobre el funcionamiento, la efectividad y la estética de cada una de ellas. En la siguiente lista se hace un breve resumen de las conclusiones obtenidas para cada una.

- **Speech Trainer**[9]

  Una herramienta gratuita que hace uso de la realidad virtual con soporte para HTC Vive y que presenta una sala oscura con una audiencia virtual a la que se enfrenta el ponente. Esta audiencia virtual está modelada en 3D con caracteres humanoides que intentan parecer realistas. Permite subir en formato PDF presentaciones propias y pasar diapositivas. Hay dos pantallas que permiten ver la presentación, una enfrente del ponente y otra detrás. El ponente tiene en una mano un micrófono y en la otra un puntero láser.

- **Presentation Simulator**[10]

  Una herramienta que, mediante realidad virtual, simula un entorno de una presentación. Tiene un coste de 10 euros, está disponible para los dispositivos Oculus y HTC Vive y requiere Steam para ejecutarse. Permite subir PDFs de presentación.

- **The Speech Improvement Company**[11]

  Esta herramienta no se ha podido probar debido a que no se ha encontrado el enlace para poder descargarla. Parece que no está disponible, pero la mecánica que ofrece es muy similar, presenta un entorno de una sala de de reuniones en la que el orador puede practicar una charla frente a avatares virtuales con forma humanoide, sentados en dicha

[9] http://speechimprovement.com/vr/

[10] http://www.presentationsimulator.com/fear-public-speaking/exposure-therapy/

[11] http://www.presentationsimulator.com/fear-public-speaking/exposure-therapy/

sala de reuniones. También tiene un entorno de un gran auditorio y una sala de conferencias.

- **Virtual Speech** [12]

  Aplicación de pago para Android y iOS disponible para diversos dispositivos de realidad virtual que permite practicar un discurso delante de un público virtual o responder a preguntas en una entrevista. También tiene un modo para practicar ventas, socialización y aprendizaje de idiomas.

- **Virtual Orator** [13]

  Virtual Orator es una herramienta para entrenar las habilidades para hablar en público, disponible para Oculus Rift, HTC Vive y Windows Mixed Reality. Permite ajustar el lugar de la presentación y el tamaño de la audiencia, así como el carácter de ésta (audiencia positiva o negativa). Tiene un coste de entre 250 y 2500 euros.

- **Speech Center VR** [14]

  Tiene un coste de 5 euros y está disponible para Samsung Gear y Daydream. Permite enfrentarse a un auditorio en el que van apareciendo distintas distracciones y el orador tiene que intentar gestionarlas correctamente.

- **BeFearless**[15]

  Producto patrocinado por Samsung que permite entrenar para hablar en público en tres modalidades: negocios, vida personal y clase. Permite conectar un smartwatch Gear S para ver las constantes vitales. El entorno, a pesar de estar creado con vídeos que se reproducen de forma repetida, da una sensación de estar plano y no genera el efecto inmersivo de tres dimensiones.

- **A Fear of Heights and other Things** [16]

  Tiene un coste de 0,99 euros y está disponible para Gear VR. Dispone de un modo para superar el miedo a hablar en público entre otros.

- **Ummo**[17]

  Herramienta que no utiliza realidad virtual, pero está enfocada a mejorar las habilidades del discurso solamente mediante la grabación de la voz de orador. Está disponible exclusivamente para iOS.

[12] `https://play.google.com/store/apps/details?id=com.virtualSpeech.android`
[13] `https://publicspeaking.tech/`
[14] `https://www.cerevrum.com/speech-center-vr`
[15] `https://www.oculus.com/experiences/gear-vr/942681562482500/`
[16] `https://store.steampowered.com/app/535460/A_Fear_Of_Heights_And_Other_Things/`
[17] `http://www.ummoapp.com/`

- **Beyond VR - Public Speaking VR**[18]

  Aplicación de Android, disponible para Google Cardboard y similares, que permite practicar presentaciones frente a un auditorio. Ofrece unas estadísticas al final sobre cómo ha ido la presentación. Los modelos del público están construidos con imágenes reales de personas y no con modelos 3D.

- **VRAVO!**[19]

  Plataforma que permite realizar presentaciones en un entorno virtual. No es un sistema para practicar, sino que sirve como medio para hacer las presentaciones online e invitar a otros usuarios a verlas, añadir elementos en 3D e interactuar con la audiencia.

- **BeChiara**[20]

  Proyecto creado para apoyar los cursos de hablar en público de la empresa Virtual Voyagers. Se utiliza la realidad virtual para presentar un entorno de entrenamiento de presentaciones al orador, y posteriormente se analizan estas presentaciones con expertos en el tema. Parece enfocado a grandes empresas.

[18] `https://play.google.com/store/apps/details?id=com.BeyondVR.beyond`
[19] `https://www.vravo.com/index_en.html`
[20] `https://www.bechiara.com/es/home`

Capítulo 3

Creación del entorno

3.1. Descripción detallada

El proyecto Hola UniVRso es un juego que pretende ofrecerle al ponente un entorno seguro para practicar sus exposiciones. Además, le ofrece realimentación a través de las reacciones de la audiencia virtual con el objetivo de que pueda mejorar sus habilidades de comunicación oral.

El sistema se apoya en dos pilares:

- Por un lado, usar la realidad virtual para crear un entorno en el que enfrentarse a una situación de hablar en público que se asemeje lo más posible a la vida real. En este entorno hay una serie de elementos básicos que el orador ve durante su inmersión en la experiencia virtual: un escenario por el que moverse libremente, una audiencia delante de él y una serie de estadísticas de su rendimiento como orador.

- Por otro lado, se hace uso de una combinación de sistemas de análisis de emociones para determinar la emoción que está transmitiendo el orador en cada momento. Estas emociones se obtienen a partir de medidas tomadas del orador en el entorno virtual, como el audio de su discurso y el contenido de éste. También se tienen en cuenta otros parámetros para identificar patrones erróneos de comportamiento, como el porcentaje de atención que dedica el orador a cada zona del público, que se mide en base a los movimientos de la cabeza detectados por el dispositivo de realidad virtual.

3.1.1. Gamificación

Uno de los mayores obstáculos que surgen a la hora de enfrentar a un ponente a su público es conseguir que éste se marque unos objetivos. En todas las aplicaciones que se han investigado antes de realizar este proyecto y que intentan ayudar a superar el miedo escénico, no hay un objetivo final en el sistema. El orador simplemente se enfrenta a una audiencia más o menos realista y puede realizar el discurso que quiera sin limitaciones de ningún tipo.

Sin embargo, en Hola UniVRso la mecánica está enfocada como un juego, dado que cuenta con una limitación de tiempo y unas estadísticas (cantidad de público atento, porcentaje de efectividad, etc.), de forma que el ponente tiene un objetivo final y puede ir modificando su discurso para lograr ese objetivo.

Hay dos bases fundamentales en los estudios de Mel Slater [50].

- El orador reacciona igual a una audiencia virtual que a una real.
- Una audiencia que da un *feedback* negativo afecta negativamente al orador y viceversa.

A pesar de eso, el objetivo de Hola UniVRso es algo que no se ha visto realizado en la literatura consultada. Hacer que la audiencia reaccione a lo que hace o siente (teniendo en cuenta sus emociones) el orador es el punto diferenciador del sistema, dado que el orador aprenderá mientras juega, y la jugabilidad estará representada en la audiencia. Es un simulador de lo que ocurre en una charla real.

Este proceso de usar mecánicas de juego en un entorno ajeno al juego se conoce como gamificación [40]. En este proyecto se ha dado gran importancia a la parte de gamificar la experiencia para conseguir mejores resultados, puesto que el orador centra sus esfuerzos en jugar, en lugar de navegar por el entorno virtual sin ningún propósito final. Además, el juego, por definición, ofrece un lugar seguro en el que, lo que ocurre dentro del juego no tiene consecuencias en el mundo real. Es el concepto de círculo mágico, acuñado por Johan Huizinga en su libro Homo Ludens [21]. Una vez dentro del círculo, se puede fallar porque sólo tendrá consecuencias dentro del círculo. Este concepto permite el fracaso, que es una de las más importantes bases del aprendizaje.

Los videojuegos son un tipo de entretenimiento muy popular en la actualidad con dos capacidades básicas: una capacidad adictiva que mantiene

enganchado al jugador durante mucho tiempo y una capacidad de enseñar al jugador casi cualquier cosa, de forma que éste pueda desarrollar habilidades útiles [51]. Precisamente estas dos capacidades hacen que los videojuegos se puedan plantear como una herramienta alternativa a la enseñanza tradicional.

Muchos investigadores aconsejan el uso de videojuegos educativos no como herramienta alternativa, sino como herramienta complementaria para mejorar el rendimiento de los estudiantes [22]. También hay estudios [10, 1] que demuestran que algunos juegos pueden ser usados para crear un entorno de aprendizaje más interesante o mejorar el interés de los alumnos y su motivación por aprender una tarea que de otra manera les resultaría muy tediosa.

Los videojuegos se pueden aplicar en muchos ámbitos de la educación: matemáticas, informática, ciencias sociales, historia o geografía. También se pueden aplicar a competencias transversales o disciplinas artísticas [30, 29, 28]. Antes los videojuegos educativos sólo se aplicaban a STEM (ciencias, tecnología, ingeniería y matemáticas, del inglés *science, tech, engineering and maths*), pero en la actualidad se aplican para educar en STEAM (*science, tech, engineering,* ***arts*** *and maths*), incluyendo las artes en la ecuación. Esto es importante puesto que hablar en público es algo artesanal que se aprende mediante la práctica, igual que otras disciplinas como la cerámica, la escultura o el teatro.

Las figuras 3.1, 3.2, 3.3 y 3.4 ilustran algunas partes del aspecto que tiene el juego.

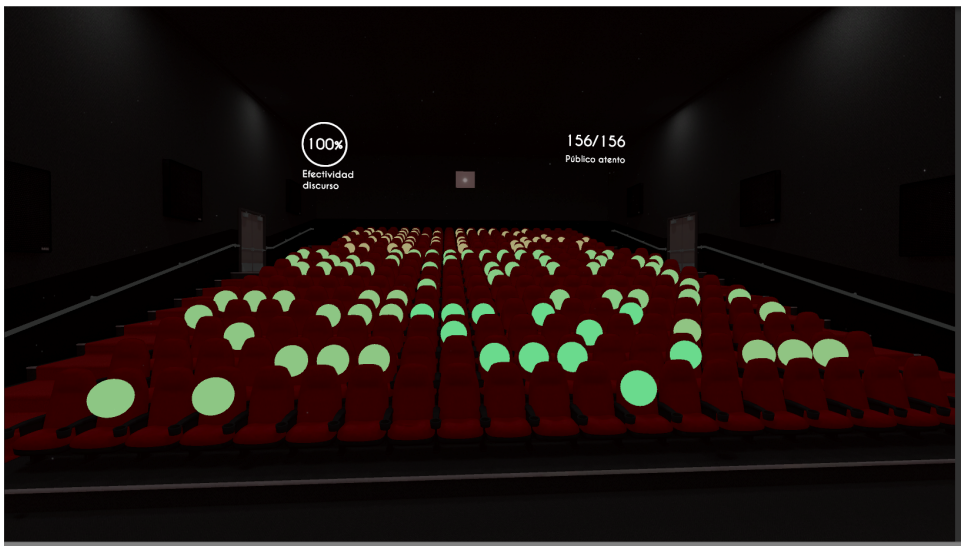

Figura 3.1: Hola UniVRso: Audiencia virtual

Figura 3.2: Hola UniVRso: Temporizador

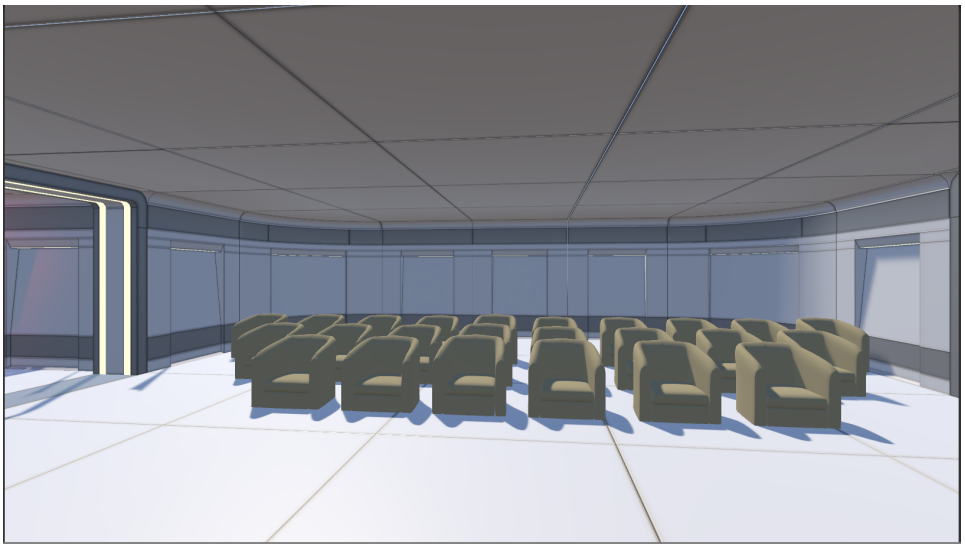

Figura 3.3: Hola UniVRso: Nivel 2 - Future

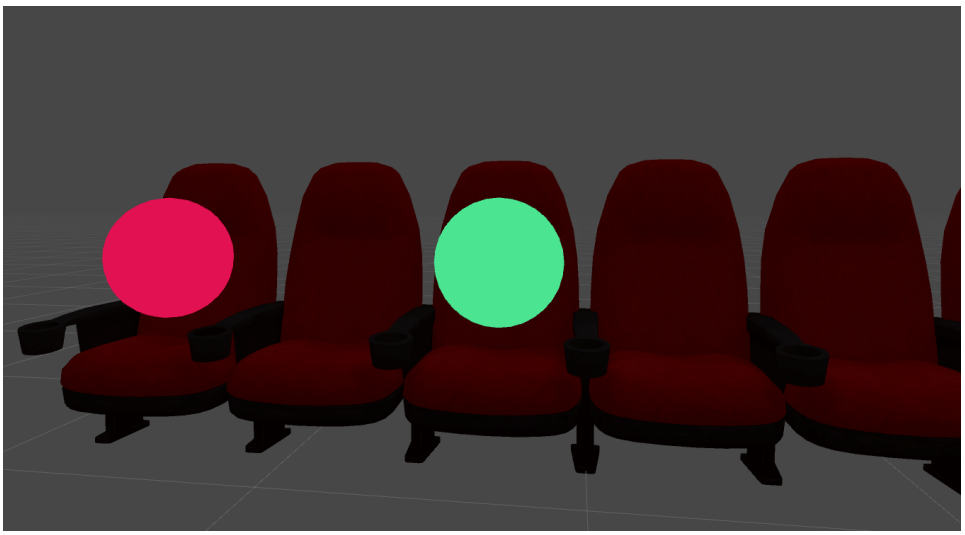

Figura 3.4: Hola UniVRso: ACMs

3.2. Arquitectura

El proyecto se divide en dos grandes subproyectos muy diferenciados. Por un lado está el entorno virtual (apartado 3.2.1), que es el sitio en el que tiene lugar la acción del orador y donde se recogen las características de la ponencia. Por otro lado, está el entorno de análisis y extracción(apartado 3.2.2), que es donde se procesan dichas características para generar reacciones en el público virtual y extraer conclusiones sobre la eficacia del orador.

3.2.1. Entorno virtual

El entorno virtual es un proyecto desarrollado con un motor de videojuegos en el que tiene la lugar la ponencia del orador. Está compuesto por un escenario sobre el que se sitúa el orador y por el que se puede mover libremente.

Frente al orador se sitúa la audiencia virtual, compuesta por un conjunto de ACMs (*Audience Character Model*) que reaccionan en tiempo real a las acciones del orador: variaciones en el discurso, zonas hacia las que dirige la mirada, tono de voz, silencios, etc.

El proyecto del entorno virtual está estructurado en escenas, en las que cada objeto 3D que hay en la escena tiene un controlador asociado. Hay un controlador general que interactúa con el resto de controladores y una serie de módulos auxiliares para realizar funciones genéricas requeridas a lo largo de todo el proceso.

A continuación se presentan las diferentes escenas (con sus respectivos subcomponentes) que conforman el entorno virtual: *Menu* (sección 3.2.1.1), Nivel 1: *Cinema* (sección 3.2.1.2) y Nivel 2: *Future* (sección 3.2.1.3), y en la sección 3.2.1.4 se mencionan algunos de los componentes auxiliares más importantes. En la figura 3.5 se pueden ver las clases principales que componen el sistema.

3.2.1.1. Menu

Escena inicial de menú para recoger los datos básicos del orador: nombre de usuario, sexo y una autovaloración de sus habilidades para hablar en público.

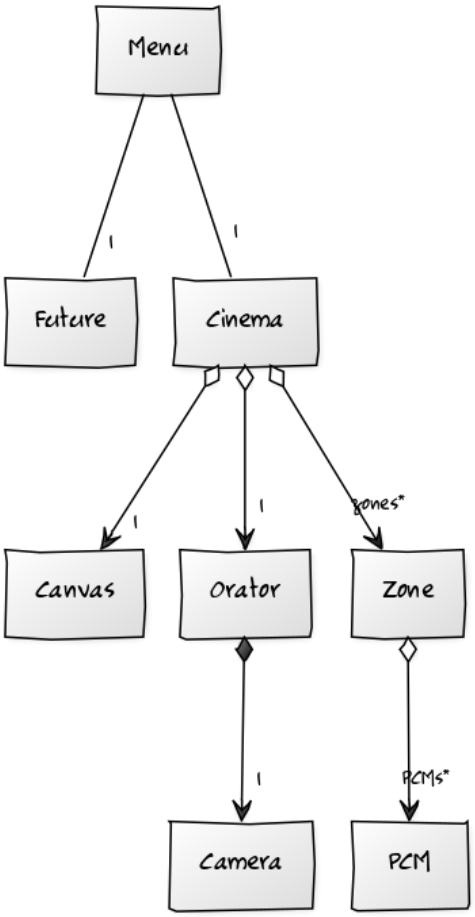

Figura 3.5: Diagrama de clases del entorno virtual

3.2.1.2. Nivel 1: Cinema

Escena principal en la que se desarrolla la sesión de juego, compuesta por un auditorio con butacas, un escenario y una pantalla. El orador se sitúa en el escenario en el entorno virtual, pudiendo moverse libremente por el espacio del escenario.

El público aparece de manera pseudoaleatoria en las butacas y cada individuo del público se denomina ACM (*Audience Character Model*). Además, en todo momento, el orador puede ver al fondo de la sala datos sobre cómo está progresando su discurso: porcentaje de efectividad, número de ACMs atentos y el estado de ánimo que está transmitiendo con la voz.

1. **Game Controller**

   Controlador principal que se activa cuando se entra en el nivel Cinema. Se encarga de controlar los estados de todos los objetos presentes en la escena, establecer la comunicación entre ellos y actualizar el estado del juego.

2. **ACM**

   Audience Character Model. Objeto en forma de esfera que representa un individuo del público. Este objeto tiene su propio controlador (`ACMControl`) que se encarga de actualizar el estado y la reacción que presenta el ACM en cada momento. Las reacciones se muestran en esta versión con variaciones del color, haciendo una **interpolación del rojo (aburrido, no atento) al verde (muy atento)**.

   El grado de atención de cada ACM se basa en el porcentaje de efectividad general del orador, que se calcula en base a los diferentes parámetros analizados en el entorno de análisis, y en base a un cierto grado de severidad que tiene cada ACM de forma individual. La severidad indica cómo de resistente es un ACM a ser cautivado por el orador. Está compuesto por varios factores, entre ellos, la posición del ACM en las butacas: a mayor distancia del escenario, mayor severidad.

   La **forma esférica** (ver figura 3.4) de los ACM se ha diseñado así dadas las limitaciones de tiempo, puesto que el modelado 3D es un proceso costoso que hubiera requerido más tiempo del que se disponía para realizar este trabajo, y porque existen estudios [55] que demuestran que los actores en un mundo virtual o en un juego no tienen por qué ser realistas. Lo importante es que dichos actores tengan un **comportamiento coherente**.

   También se ha tomado esta decisión por una cuestión de simplicidad, ya que las esferas son un objeto homogéneo que, a pesar de no tener forma humanoide, en el entorno de juego puede representar perfectamente a una persona de una audiencia real, debido al efecto inmersivo, que provoca que el cerebro dé por válidas ciertas premisas que no se cumplen en el mundo real. Siguiendo ese principio de simplicidad, las reacciones son también muy sencillas de captar a simple vista: el color general del público o de una zona del público puede dar una idea clara al orador sobre cómo está progresando su presentación y si debe variar su discurso, modificar su tono de voz o mirar más hacia una parte del público que tiene olvidada.

3. **Zones**

   Las zonas son objetos invisibles superpuestos sobre diferentes áreas del público cuya función es determinar el grado de atención que está prestando el orador a ese área mediante su controlador: `ZoneControl`.

Las zonas engloban a un número determinado de ACMs y éstos basan en parte su estado de atención en la frecuencia con la que está mirando el orador a la zona en la que se encuentran.

4. **Orator**

   El objeto del orador representa al orador dentro del entorno virtual y contiene la cámara que representa los ojos del ponente, que determina lo que está viendo en cada momento y afecta al estado general del juego y a las reacciones de la audiencia. Por el momento, el objeto no tiene un cuerpo físico definido en 3D, por la misma razón de simplicidad y tiempo que se da para el objeto ACM. Este objeto tiene su propio controlador (`OratorControl`) que se encarga de actualizar el estado del orador, de su cámara y de su posición en el escenario.

5. **Canvas**

   El canvas es un objeto que hace las funciones de HUD (*Head-up Display* o barra de estado) que se encuentra al fondo de la sala en el entorno virtual y que, al estilo de cualquier juego, muestra al orador datos más detallados sobre su sesión de juego: el número de ACMs que están atentos, el porcentaje de efectividad de su discurso y también la emoción que está transmitiendo a través de la voz en cada momento. En cada uno de los laterales de la sala el orador puede ver un cronómetro que le indica el tiempo de sesión transcurrido y cuánto queda hasta el final.

3.2.1.3. Nivel 2: Future

La escena Future se corresponde con otro nivel en el que puede desarrollarse la sesión de juego, contando éste con un auditorio más pequeño con butacas, un escenario y una pantalla. El orador se sitúa en el escenario en el entorno virtual, pudiendo moverse libremente en el espacio del escenario.

Tiene el mismo funcionamiento que la escena Cinema, con la salvedad de que el entorno es distinto y tiene lugar en un vestíbulo más iluminado, con menos público y con un tono más informal. El nombre de la escena surge debido a que los materiales utilizados para construir el entorno pertenecen a un conjunto de recursos de una nave espacial futurista (ver figura 3.3). Esta escena se ha preparado para ofrecer al orador otro nivel en el que el entorno varía, dando la posibilidad de jugar con la misma mecánica en un escenario distinto.

3.2.1.4. Auxiliares

Además de las escenas, existen otros elementos auxiliares en el proyecto que tienen mucha importancia para el correcto funcionamiento de la arquitectura.

1. **Request Helper**

   `RequestHelper` es un componente auxiliar que actúa como **puente** entre el entorno virtual y el entorno de análisis. Su objetivo principal es transmitir los datos recogidos (audio del orador y hacia dónde mira) al entorno de análisis para su posterior procesamiento. También recibe las reacciones en tiempo real, que van llegando desde el entorno de análisis a partir de esos datos generados, para transmitirlas al control del juego, que se encarga de actualizar las reacciones de la audiencia.

2. **Character Motor**

   Controlador del orador que se ha utilizado para poder hacer las pruebas con el ratón y el teclado, sin necesidad de estar usando las gafas de realidad virtual de forma continua. El funcionamiento es muy similar a cualquier juego, permitiendo utilizar las teclas `ASDW` para mover al orador sobre el escenario y el ratón para desplazar su mirada. Este elemento auxiliar ha resultado muy valioso para ahorrar tiempo en la fase de pruebas de la cámara del orador, para determinar hacia dónde está mirando y generar las características adecuadas para enviar al entorno de análisis.

3.2.2. Entorno de análisis

El proyecto del entorno de análisis es una API[1] REST que se encarga de **procesar las características o rasgos de la ponencia que recibe del entorno virtual**. Una vez procesados y analizados estos rasgos, el sistema genera un porcentaje de efectividad del orador, a través de las emociones extraídas de las características y de otros análisis. Este porcentaje de efectividad se traduce posteriormente en reacciones de la audiencia virtual, que proporciona realimentación al orador en tiempo real sobre cómo está siendo su discurso, y también forma parte de un compendio de datos que se procesan al finalizar la sesión para generar un resultado final más detallado sobre toda la sesión.

[1] `http://www.ticbeat.com/tecnologias/que-es-una-api-para-que-sirve/`

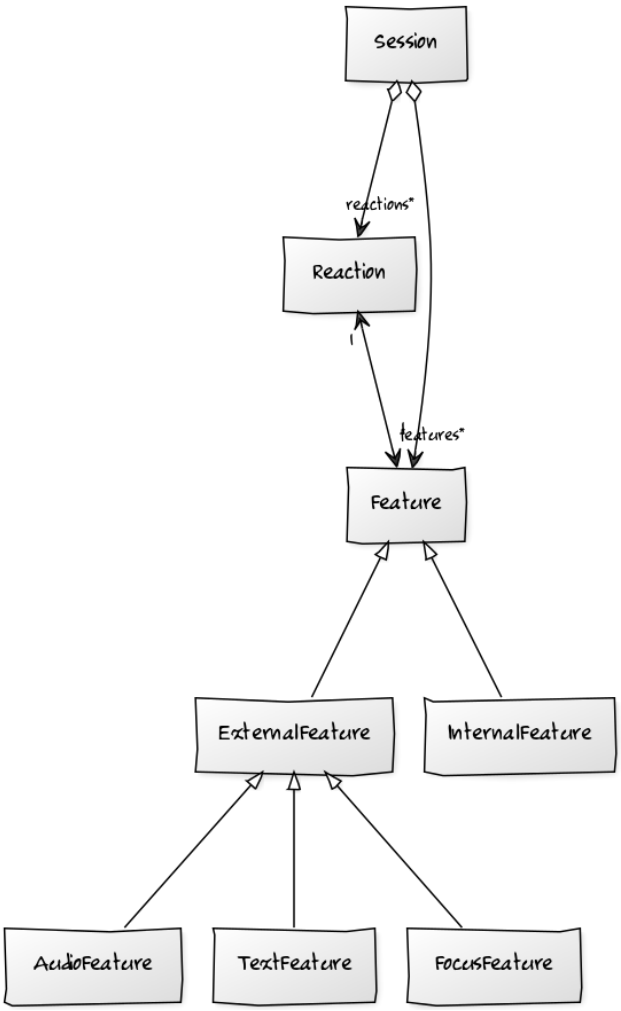

Figura 3.6: Diagrama de clases del entorno de análisis

El proyecto se organiza en cuatro grandes apartados: modelos (sección 3.2.2.1), controladores (sección 3.2.2.2), extractores (sección 3.2.2.3) y auxiliares (sección 3.2.2.4). En la figura 3.6 se puede ver un diagrama de las clases principales del entorno de análisis.

3.2.2.1. Modelos

Los modelos o clases son componentes en los que se estructuran las diferentes entidades con las que trabaja el sistema.

1. **Session**

   En el momento en que comienza el juego y se rellenan los datos del ponente, se crea una nueva sesión. Todos los datos guardados y analizados, así como las reacciones generadas están ligadas a dicha sesión. Este modelo engloba toda la experiencia de juego para cada jugador, y además sirve para hacer un análisis posterior más detallado de lo que ha ido ocurriendo durante la sesión de juego.

2. **Feature**

   Feature o característica hace referencia a datos recogidos a partir de las acciones del orador en el entorno virtual. Es uno de los modelos esenciales de este proyecto, ya que almacena la información que se analiza y de la que se extraen las reacciones a posteriori. Se pueden distinguir dos grandes subtipos:

   a) **Internal Feature**

   Características internas del orador. Hacen referencia a aquellos parámetros del orador que el público no ve y que están más ligados a **medidas biométricas del orador**: por ejemplo, sudoración o ritmo cardíaco. Estas características no tienen por qué afectar a la reacción de la audiencia, puesto que ésta no es consciente de ellas, pero sí son interesantes para saber cómo afecta el estado interno del orador a sus acciones y a su desempeño como ponente.

   b) **External Feature**

   Características externas del orador. Son aquellas **características claramente detectables por la audiencia** y que, por lo tanto, van a tener un impacto en el grado de atención del público. Se pueden distinguir entre distintos tipos de características según el parámetro que se esté analizando en cada momento:

   1) **Audio Feature**

   Característica de audio. Se trata de un fragmento de audio de corta duración (5 - 10 segundos) grabado del discurso que está haciendo el orador. Este tipo de características se graban de forma superpuesta, con una ventana de grabación (se procesan 10 segundos iniciales y después se genera un análisis cada 5 segundos), y servirán posteriormente para analizar el tono de voz, sin tener en cuenta el contenido del discurso, y detectar emociones del orador a partir de dichos fragmentos.

   2) **Text Feature**

   Característica de texto. Este tipo de característica sí que se centra en el contenido del discurso y se obtiene a partir de los fragmentos de audio mencionados anteriormente. Se procesan estos fragmentos para convertir el audio en el texto que genera

este tipo de *features*. Posteriormente, serán analizadas para determinar la emoción del orador según el contenido de su discurso.

3) **Focus Feature**
Característica de atención del orador. Este tipo de característica hace referencia al porcentaje de atención que está prestando el orador a cada una de las zonas del auditorio. Aquí no se pretende analizar la emoción que está transmitiendo el orador, sino verificar si está mirando por igual a toda la audiencia, está focalizado en una zona del público o no deja de mirar a la pantalla de las diapositivas, sin prestar atención al público.

3. **Reaction**

   La reacción es un modelo que se encarga de almacenar la reacción generada a partir de una característica. Contiene la emoción extraída de una característica y el porcentaje de efectividad del discurso que se calcula en base a dicha emoción. También hace referencia a la característica de la que se ha obtenido para poder hacer posteriores análisis y comparaciones.

3.2.2.2. Controladores

Los controladores contienen los métodos asociados a las funciones principales de la API, a los que se podrán realizar peticiones para actuar sobre los modelos anteriormente mencionados.

1. **Feature Controller**

   Controlador de características. Este controlador agrupa los métodos para crear objetos del tipo Feature y sus subtipos. Se utiliza principalmente para recibir peticiones de creación de características del orador para almacenarlas en la base de datos.

2. **Reaction Controller**

   Controlador de reacciones. Este controlador agrupa los métodos que permiten consultar las reacciones que se han generado en los extractores, a partir de las características analizadas del orador. Se utiliza principalmente para recibir peticiones de consulta de reacciones pendientes de enviar a la audiencia virtual.

3. **Session Controller**

   Controlador de sesión. Este controlador agrupa los métodos para crear y modificar las sesiones. Se utiliza principalmente para recibir peticiones de gestión de sesión: crear una sesión, dar comienzo al juego o finalizar una sesión.

3.2.2.3. Extractores

Los extractores forman el núcleo del proyecto de análisis. En ellos es donde se hace el procesamiento principal de la API, dado que reciben las características del orador, las analizan, extraen las emociones presentes en dichas características y, por último, modifican la atención de la audiencia en tiempo real mediante un algoritmo de generación de reacciones.

Sin embargo, antes de proceder con el análisis de emociones, es necesario hacer un tratamiento de los datos para su correcto análisis.

Tratamiento del audio

Para hacer el análisis de emociones a partir del audio es necesario tratar esos segmentos de sonido antes de enviarlos al análisis. No se puede enviar el sonido en bruto, dado que afecta en gran medida al funcionamiento del extractor y a la corrección de resultados. Por ello, se convierten los segmentos de audio del formato de grabación a formato `WAV`, con 8KHz de tasa de muestreo, 16 bits de profundidad (16 bits en cada muestra) en un único canal (no estéreo).

Este tratamiento se hace directamente en el entorno de análisis, dado que los sistemas de grabación difieren de un entorno a otro, y por ello se relega esta funcionalidad a la API, para que el cliente pueda enviar el audio en bruto sin preocuparse del formato.

También es importante que la calidad de la grabación sea buena. Por ello, no se recomienda usar dispositivos de grabación con codificadores que hagan una gran compresión del audio como los que se usan para Voz sobre IP.

Las grabaciones de poca calidad contienen deformidades, silbido excesivo y otros contaminantes vocales. Dichas grabaciones no son apropiadas para el análisis. Por lo general, es recomendable usar un dispositivo de grabación de buena calidad o, en su defecto, un micrófono de un teléfono móvil ya que suele generar una calidad de grabación bastante superior a la de un ordenador portátil, por ejemplo.

Además, el sonido debe contener la grabación de un único discurso, es decir, solamente debe estar hablando una persona para un análisis correcto. Por ello, a pesar de que los sistemas de análisis de emociones tienen cierta tolerancia al ruido, dado que se hace otro procesamiento de la señal acústica antes de analizarla, hay que evitar grabar con ruido de fondo y procurar hacerlo en una estancia con ruido relativamente bajo.

Finalmente, esa señal de audio tratada puede guardarse como `AudioFeature` para hacer la extracción de emociones a partir de la señal de la voz, o puede sufrir un segundo tratamiento para convertirla en texto, proceso que se hace mediante una librería de conversión del discurso a texto. Este paso final convertiría la característica al tipo `TextFeature` y serviría para poder hacer el análisis de emociones a partir del contenido del discurso.

De momento, existen tres tipos de extractores en el proyecto. Se ha elegido analizar tres tipos de variables para este primer prototipo. Se hace un análisis del audio sin tener en cuenta el contenido, debido a que en la lecturas consideradas en este proyecto se posiciona como uno de los elementos más importantes para dar un buen discurso, hecho que se demostrará cierto en el apartado de experimentos. Por otro lado, se analiza el contenido de la charla puesto que aquello que se dice confiere emociones al cómo se dice, y estas emociones pueden afectar al público.

Por último, se hace un análisis de hacia dónde mira el orador, si presta más atención a una parte de la audiencia que a otra, o si está continuamente mirando la pantalla. Esta variable permite extraer la reacciones a otro componente esencial a la hora de dar una charla: contacto visual con la audiencia.

1. **Audio Extractor**

   Extractor de audio. Contiene métodos de análisis de objetos de tipo `AudioFeature` y permite extraer emociones a partir de esas características, usando diferentes librerías de extracción de emociones a partir de la voz.

2. **Text Extractor**

   Extractor de texto. Contiene métodos de análisis de objetos de tipo `TextFeature` y permite extraer emociones a partir de esas características de texto, centrándose en el contenido del discurso y no en el sonido, mediante el uso de diferentes librerías de extracción de emociones a partir del texto.

3. **Focus Extractor**

   Extractor de reacciones a partir de hacia dónde dirige el orador la mirada (`FocusFeature`). Para generar las reacciones tiene en cuenta el porcentaje de atención que dedica el orador a cada zona del público y a la pantalla del proyector.

3.2.2.4. Auxiliares

Los auxiliares engloban a todos aquellos métodos que se utilizan en varios sitios del proyecto o bien aquellos cuya funcionalidad no se puede encuadrar en ninguno de los otros apartados. Cabe destacar:

1. **Speech to text Helper**

   Auxiliar que permite hacer la transformación de los segmentos de audio en texto para poder generar objetos del tipo `TextFeature`, que posteriormente se procesarán en el `TextExtractor`.

2. **Audio Helper**

   Auxiliar que engloba todos los métodos de de tratamiento del audio para convertirlo al formato adecuado para el procesamiento.

El sistema se ha planteado de una forma modular para permitir la inserción de componentes de extracción o análisis de forma sencilla, requiriendo la mínima configuración. La implementación se ha hecho teniendo en mente siempre la escalabilidad del proyecto. Por ejemplo, es posible crear en el sistema nuevos tipos de características que nos interese guardar, simplemente añadiendo un nuevo modelo `XFeature` con nuevos atributos, que herede del modelo padre Feature o, en su caso, `ExternalFeature` o `InternalFeature`. Así mismo, se pueden añadir uno o varios extractores que permitan analizar dicha característica, generando reacciones que se añadirán a la suma de reacciones que se mandan al entorno virtual, mejorando las reacciones de la audiencia virtual y dando una realimentación más realista al orador.

Además de esto, el proyecto cuenta con una configuración de rutas para exponer las rutas de la API, a las que posteriormente hará peticiones el entorno virtual o cualquier otro entorno que desee interactuar con la API.

Las rutas están creadas siguiendo el modelo REST (*REpresentational State Transfer*), que es un protocolo software cliente-servidor sin estado, donde ni el servidor ni el cliente necesitan guardar información sobre el estado del sistema, sino que las peticiones son autocontenidas y tienen toda la información necesaria para realizar el procesamiento.

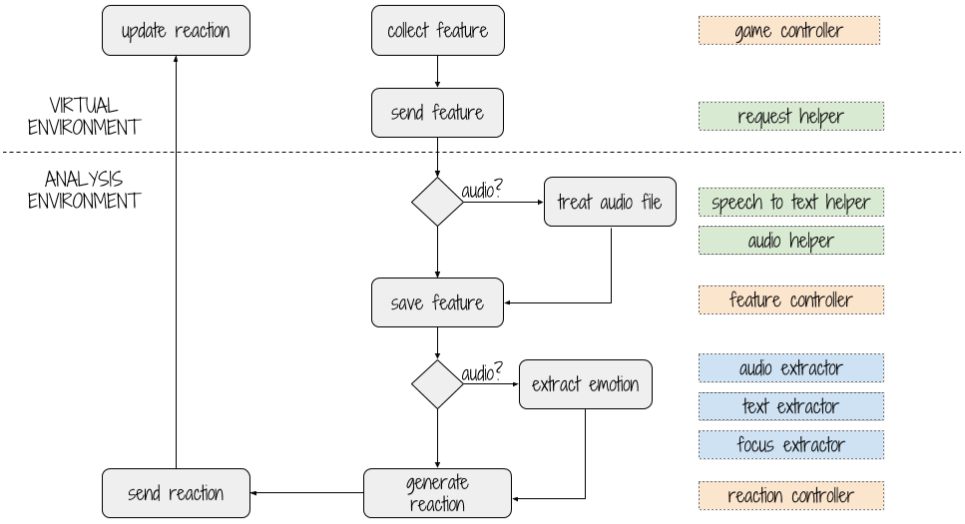

Figura 3.7: Diagrama de flujo de Hola UniVRso

Las operaciones más importantes en los sistemas REST y la especificación `HTTP` son: `POST` (crear), `GET` (consultar), `PUT` (modificar) y `DELETE` (borrar).

Las grandes ventajas de los sistemas REST es que separan cliente y servidor, proporcionan una interfaz estándar para hacer los diferentes tipos de peticiones (crear, modificar, borrar) y son independientes del lenguaje. Esta última ventaja es muy importante puesto que permite hacer peticiones desde cualquier sistema, sin importar el lenguaje de programación.

En la figura 3.7 se puede ver un diagrama del funcionamiento general de los dos entornos, desde que se recoge un parámetro del orador en un determinado momento hasta que se genera una reacción que se envía al entorno virtual, así como los componentes involucrados en cada proceso.

Figura 3.8: Dispositivo HTC Vive

3.3. Tecnologías

En este apartado se describen las tecnologías y herramientas utilizadas para construir los dos subproyectos que conforman Hola UniVRso.

3.3.1. Entorno virtual

Algunas de las tecnologías en el mercado más importantes que generan entornos de realidad virtual son, entre otras, HTC Vive, Oculus Rift, Playstation VR, Samsung Gear y Google Cardboard.

3.3.1.1. Tecnologías disponibles

- **HTC Vive**[2]

  HTC Vive es un dispositivo de realidad virtual desarrollado por HTC y Valve Corporation. Este dispositivo usa sensores colocados en dos o más puntos de una habitación que permiten seguir los movimientos del usuario y reproducirlos en el entorno 3D. También dispone de controladores de mano que permiten interactuar con el entorno. La versión Pro ha salido recientemente al mercado y tiene un precio de $799.

[2] `https://www.vive.com/`

Figura 3.9: Dispositivo Oculus Rift

Figura 3.10: Dispositivo PlayStation VR

- **Oculus Rift**[3]

  Oculus Rift es un casco de realidad virtual desarrollado por Oculus VR. Se lanzó al mercado en abril de 2016. Se le pueden incluir controladores manuales para simular las manos. También permite añadir estaciones base para mapear la posición del usuario en la estancia. Su precio actual es de 449 euros.

- **Playstation VR**[4]

  PlayStation VR es un visor de realidad virtual desarrollado por Sony Interactive Entertainment que salió a la venta el 13 de octubre de 2016. Está diseñado específicamente para la consola PlayStation 4. Este dispositivo puede dar salida a una imagen para el visor PlayStation VR y un televisor al mismo tiempo, con la televisión ya sea reflejando la imagen que aparece en el visor, o mostrando una imagen separada para el juego competitivo o cooperativo. El precio actual de este dispositivo es de alrededor de 300 euros.

- **Samsung Gear VR**[5]

  Samsung Gear VR es un dispositivo de realidad virtual desarrollado por Oculus que permite transformar un smartphone de Samsung en

[3] `https://www.oculus.com/`
[4] `https://www.playstation.com/es-es/explore/playstation-vr/`
[5] `https://www.samsung.com/global/galaxy/gear-vr/`

Figura 3.11: Dispositivo Samsung Gear VR

Figura 3.12: Dispositivo Google Cardboard

un dispositivo de realidad virtual portátil. La diferencia principal de este dispositivo es la necesidad de disponer de un móvil de Samsung. Sin embargo, no necesita de nada más para operar y crear el entorno virtual. En la actualidad tiene un precio de $130.

- **Google Cardboard**[6]

  Google Cardboard es una plataforma de realidad virtual desarrollada por Google sobre una base de cartón plegable, de ahí su nombre. El funcionamiento consiste en montar el dispositivo en cartón siguiendo las instrucciones o comprarlo, e instalar la aplicación de Google que permite duplicar la pantallas para presentarlo en modo de realidad virtual en el teléfono móvil sobre el que se vaya a probar. El coste de este dispositivo es variable: puede ser gratuito en el caso de hacerlo en casa o se puede comprar a diferentes precios, dependiendo de los modelos y materiales.

[6] `https://vr.google.com/cardboard/`

Figura 3.13: Dispositivo Windows Mixed Reality

- **Windows Mixed Reality**[7]

  Windows Mixed reality es un dispositivo que combina realidad virtual con realidad aumentada en una única plataforma. El funcionamiento es similar a los dispositivos mencionados anteriormente. Cuenta con un HMD que permite entrar en el mundo virtual y permite utilizar controladores manuales también. Tiene un coste de $299.

3.3.1.2. Tecnologías seleccionadas

Para el entorno virtual, se ha seleccionado la plataforma de **Unity 3D**[8], un motor de videojuegos multiplataforma creado por Unity Technologies, para llevar a cabo el desarrollo. Esta decisión se ha tomado por dos motivos principales:

1. La facilidad de uso, ya que permite trabajar con entornos 3D de forma sencilla e importar contenido ya hecho para un desarrollo más ágil

2. Y también porque integra cualquiera de los dispositivos de realidad virtual que se mencionan en este apartado sin necesidad de implementar los controladores para dichos dispositivos.

Además, es una plataforma gratuita, con una gran comunidad online que proporciona soporte a los diferentes problemas que puedan surgir, y un amplio catálogo de recursos gráficos de los cuales muchos son de uso gratuito, algo que ha resultado imprescindible en este trabajo para realizar el prototipado y poder iterar con rapidez probando diferentes entornos.

[7] `https://www.microsoft.com/en-us/store/collections/vrandmixedrealityheadsets`

[8] `https://unity3d.com/es`

Se ha seleccionado el dispositivo de **HTC Vive** para desarrollar el proyecto, dada su facilidad para integrarlo en el entorno de programación, sin necesidad de realizar implementación de controladores del dispositivo, y también debido a su disponibilidad para su uso. Sin embargo, cualquiera de los dispositivos que se han enumerado anteriormente sería perfectamente válido y podría usarse para simular el entorno de Hola UniVRso.

3.3.2. Entorno de análisis

En el entorno de análisis se ha hecho uso de diferentes tecnologías para hacer la extracción de emociones a partir de las características del orador. Aunque finalmente no se han incluido características internas en el prototipo actual, sí que se han investigado y se han hecho pruebas con tecnologías para medir parámetros internos o biométricos del orador. Entre estas herramientas cabe destacar el dispositivo de MySignals. También se han probado múltiples herramientas para el análisis de emociones en la voz: BeyondVerbal, Vokaturi, Good Vibrations, Pitch JS y OpenSMILE entre otras. Para el análisis de texto se han investigado las herramientas de IBM Watson Tone Analyzer API y Bitext API.

3.3.2.1. Tecnologías disponibles

- **MySignals**[9]

  MySignals es un dispositivo, creado por la empresa española Libelium, que consiste en una plataforma de medición de constantes médicas para crear aplicaciones relacionadas con salud o para crear nuevos dispositivos médicos de medición a partir de la base de MySignals. Contiene 17 sensores que permiten medir más de 20 parámetros biométricos: ritmo cardíaco, ECG (Electrocardiograma), temperatura corporal o presión arterial entre otros. Ofrece una API para interactuar con los datos recibidos por los sensores y también permite subir los datos directamente a su propia plataforma en la nube.

- **BeyondVerbal**[10]

  BeyondVerbal es una herramienta desarrollada por la compañía Beyond Verbal Communication que ofrece una API online que hace análisis de emociones a partir de las variaciones del tono de la voz sin tener en cuenta el contenido del discurso. En la propia compañía usan su herramienta para extraer marcadores en la voz de los individuos que per-

[9] `http://www.libelium.com/`
[10] `http://www.beyondverbal.com/`

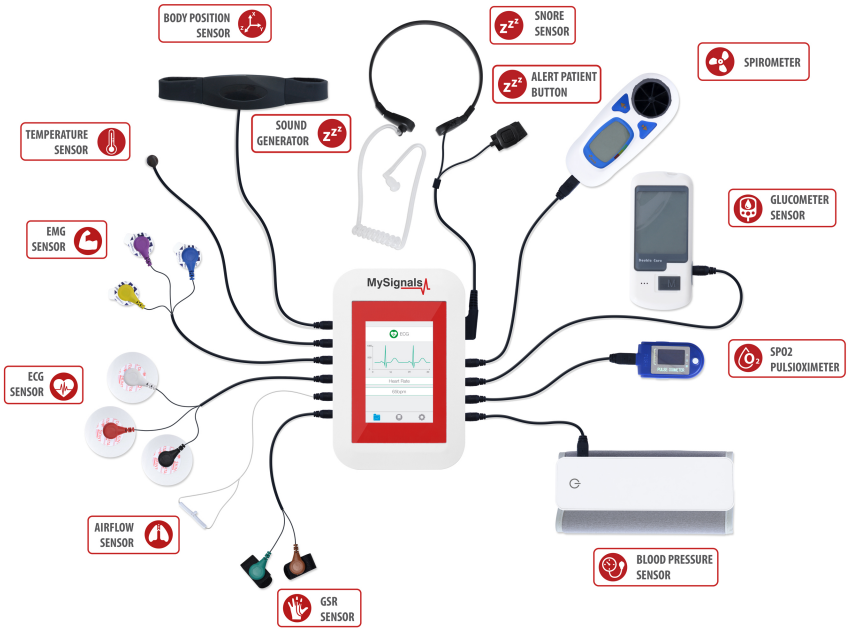

Figura 3.14: Dispositivo MySignals

mitan detectar ciertas enfermedades como depresión [35], esquizofrenia [32], autismo [5] o incluso la enfermedad coronaria [31]. También se usa para detectar el estado emocional del individuo en ciertas situaciones en las que este estado emocional puede tener un efecto negativo(por ejemplo, conduciendo un coche).

- **Vokaturi**[11]

  Vokaturi es una herramienta software que hace reconocimiento de emociones en la voz humana. Sus algoritmos los ha diseñado Paula Boersma [4], profesora en la Universidad de Ámsterdam. La herramienta puede determinar, directamente a partir de la voz, emociones como felicidad, tristeza, miedo, enfado o estado neutral. Dispone de varias versiones, una de ellas de código abierto y gratuita para uso no comercial, aunque esta última garantiza un porcentaje más bajo de acierto que las versiones premium. La herramienta está programada en C y dispone de varias librerías cliente en diferentes lenguajes para su uso. Para poder utilizarla es necesario descargar la librería e importarla en el código.

- **Good Vibrations**[12]

[11] `https://vokaturi.com/`
[12] `http://www.good-vibrations.nl/api`

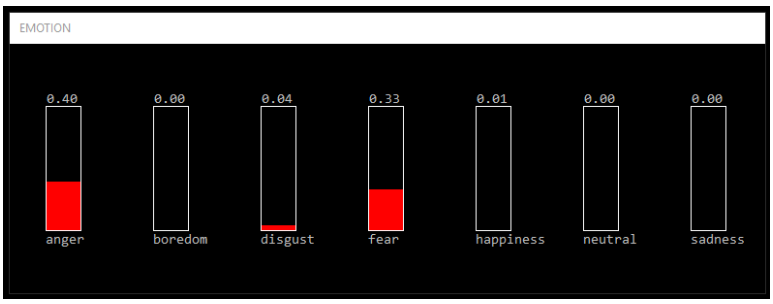

Figura 3.15: Emovoice: Análisis de segmentos de audio

Good Vibrations API es una herramienta, desarrollada por The Good Vibrations Company, cuyo objetivo es mejorar el rendimiento de los individuos mediante el análisis de señales biológicas y otro tipo de señales. Se centran en el análisis de señales proporcionadas por la voz de sujeto, el corazón, los pulmones y otros órganos. Analizando estas señales infieren aspectos relacionados con la salud, el estado emocional o la excitación experimentada por un individuo. Se han hecho diferentes pruebas con la API de la que disponen para analizar audio de voz. Para poder realizar pruebas con esta herramienta es necesario descargar su SDK e importarlo en el proyecto. Disponen de varias librerías cliente para poder enlazarla con otros lenguajes de programación.

- **Pitch JS**[13]

  Pitch JS es una librería de código abierto desarrollada en Javascript para analizar las variaciones en el tono de voz de un audio. Muestra estadísticas de dichos cambios y la variación del tono, vibración y volumen de la voz. Sin embargo, no incluye análisis de emociones a partir de esos datos.

- **EmoVoice**[14]

  Emovoice es un proyecto de investigación [57, 58], liderado por la profesora Elisabeth André y el Dr. Johannes André de la Universidad de Augsburgo, que consiste en un framework desarrollado para el reconocimiento en tiempo real de emociones a partir de propiedades del discurso, sin tener en cuenta el contenido. El proyecto está desarrollado en Python y es de código abierto. Para poder usarlo es necesario descargarlo e importarlo. En el proyecto vienen además incluidos múltiples ejemplos para probar la herramienta y validar su efectividad.

  El problema principal que presenta este proyecto y por el que se ha

[13] `https://github.com/audiocogs/pitch.js/`
[14] `https://github.com/hcmlab/emovoice`

descartado como tecnología a usar es que la base de datos [13] que han usado para entrenar el modelo de predicción es un conjunto de discursos en alemán. Puesto que este proyecto está planteado, por el momento, para tratar discursos en castellano y hay algunas diferencias notables en la expresión de emociones en el discurso según el idioma hablado, las predicciones del sistema no eran acertadas para discursos en castellano.

- **OpenSMILE**[15]

  El proyecto [13] OpenSMILE (*Open Speech & Music Interpretation by Large-space Extraction*) empezó en la Universidad Técnica de Múnich en 2018 en el marco de un proyecto europeo, liderado por Florian Eyben, Martin Wöllmer, and Björn Schuller. Es una herramienta que permite extraer características de grandes archivos de audio en tiempo real. Sirve tanto para el procesamiento de discurso como para piezas de música. Permite detectar emociones en un discurso, analizar interacciones sociales, detectar patrones de estrés, etc. Para poder utilizarla basta con descargar un binario o utilizar la API disponible de forma gratuita para uso personal o de investigación. En este caso ocurre lo mismo que con la herramienta anterior: está entrenada con bases de datos en otros idiomas que se diferencian mucho en cuanto a las emociones que transmiten en el discurso.

- **IBM Watson Tone Analyzer API**[16]

  Tone Analyzer API es una de las herramientas disponibles en el framework de análisis de IBM Watson [20]. Es una herramienta que consiste en una API que analiza emociones a partir de texto de cualquier tipo. Dado que las personas muestran distintos tonos, como alegría, tristeza, ira y simpatía en sus comunicaciones diarias, estos tonos o emociones pueden repercutir en la eficacia de la comunicación en distintos contextos.

- **Bitext API**[17]

  Bitext es una herramienta de pago de análisis de texto, que extrae emociones a partir del texto. Está disponible en más de 50 idiomas y genera valoraciones del grado de positividad o negatividad del texto, tal como se puede ver en la figura 3.17.

Figura 3.16: IBM Watson Tone Analyzer API: Análisis de texto

Figura 3.17: Bitext API: Análisis de texto

3.3.2.2. Tecnologías seleccionadas

El entorno de análisis se ha construido utilizando el framework de desarrollo **Node.js**[18], en el que se desarrolla con el lenguaje de programación **Javascript**. Este framework es uno de los entornos de ejecución de Javascript más usados para servidor. Node.js utiliza un modelo de operaciones Entrada/Salida sin bloqueo y orientado a eventos, de forma que el sistema se mantiene ligero y eficiente. Cuenta con un sistema de paquetes denominado `npm`, que es el ecosistema de librerías de código abierto más grande del mundo. En la figura 3.18 se puede ver una comparativa de este framework con otros.

[15] `https://audeering.com/technology/opensmile/`
[16] `https://www.ibm.com/watson/services/tone-analyzer/`
[17] `https://www.bitext.com/`
[18] `https://nodejs.org/es/`

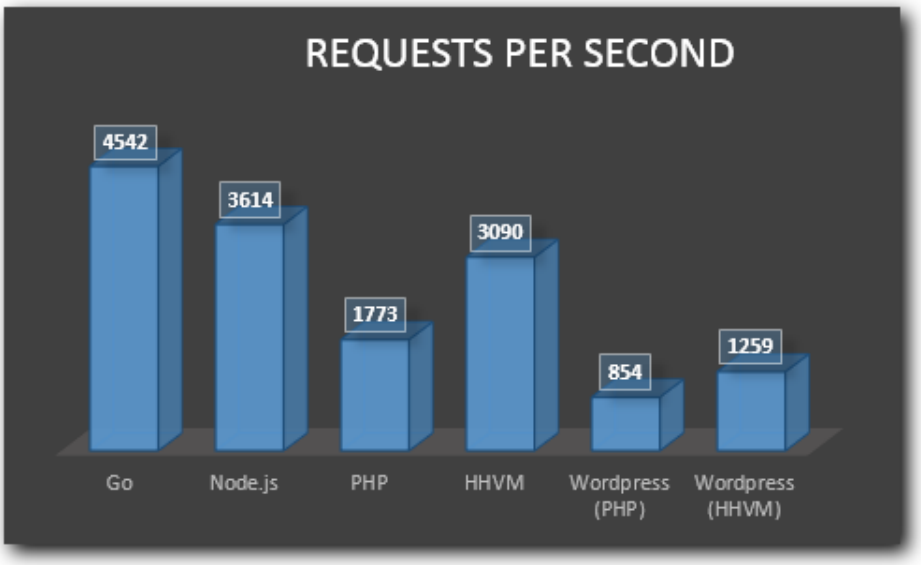

Figura 3.18: Comparativa de peticiones por segundo de Node.js en comparación con otros frameworks

Node.js está orientado a **eventos asíncronos**, es decir, un modelo pensado para escenarios en los que el código de usuario deba ejecutarse al producirse un evento, en lugar de una función detrás de otra como en otros lenguajes como Java o C. El sistema se basa en callbacks, que son aquellas funciones que se ejecutan cuando se lanza un evento. En el siguiente ejemplo se puede ver una pequeña aplicación de servidor creada con Node.js.

```
const http = require('http');

const hostname = '127.0.0.1';
const port = 3000;

const server = http.createServer((req, res) => {
  res.statusCode = 200;
  res.setHeader('Content-Type', 'text/plain');
  res.end('\textsc{Hola UniVRso}\n');
});

server.listen(port, hostname, () => {
  //callback
  console.log('El servidor se está ejecutando');
});
```

Figura 3.19: MongoDB: Estructura de documentos

En el ejemplo anterior, la función de callback será ejecutada para cada conexión. Si no hay actividad, el servidor de Node.js se queda en estado de suspensión.

Este modelo de funcionamiento resulta ideal para el trabajo que se plantea en este documento, puesto que se puede tener un servidor con el entorno de análisis en el sistema local o desplegado y distribuido en la Nube. Ese servidor se queda latente a la espera de peticiones por parte del entorno virtual y, una vez que recibe esas peticiones, lanza una cadena de eventos que, a su vez, ejecutan callbacks a medida que se van completando procesos en el sistema, desde guardar una característica del orador en la base de datos hasta generar una reacción cuando se completa el análisis emocional de un segmento de la presentación.

Además, la inmensa variedad de librerías de código abierto disponibles para importar en cualquier proyecto de Node.js hace que montar un proyecto con funcionalidades básicas sea un proceso sencillo y rápido.

Para guardar los datos se ha optado por **MongoDB**[19], una base de datos NoSQL [2, 6, 33] de código abierto estructurada en documentos. En lugar de guardar los datos en tablas como las bases de datos al uso, MongoDB almacena los datos en forma de documentos con una estructura similar a objetos `JSON`. La gran ventaja de MongoDB es que proporciona un esquema dinámico, de forma que los atributos de los objetos pueden variar en el tiempo y no tienen una estructura rígida.

MongoDB está pensado además para guardar grandes cantidades de datos y tiene distribución y replicación integrada, de modo que es relativamente sencillo replicar y distribuir los datos para escalar el sistema.

La herramienta de análisis de voz seleccionada es **BeyondVerbal**, cu-

[19] `https://www.mongodb.com`

Figura 3.20: BeyondVerbal: Valores de Temper

yo funcionamiento es simple: todo el procesamiento se hace en su entorno Cloud. Desde el cliente es necesario mandar peticiones separadas por intervalos con los segmentos de audio a analizar. La respuesta que devuelve la API contiene valores de la **valencia, temperamento y excitación de la voz** analizada, así como el **grupo de emociones** en el que se ha clasificado (Enfado/Disgusto/Estrés, Tristeza/Inseguridad/Aburrimiento, Neutral, Felicidad/Entusiasmo/Simpatía o Calidez/Calma).

Esta herramienta tiene una versión gratuita, que tiene una limitación de peticiones al mes, y tiene otras versiones de pago cuyo precio difiere en función del número de peticiones y de si ofrecen o no soporte.

Se ha seleccionado esta herramienta debido a su fácil integración, a que dispone de versión gratuita y a que las pruebas que se han hecho en los experimentos que se detallan en apartados posteriores han tenido una buena tasa de acierto.

Los resultados que devuelve la API están agrupados en diferentes parámetros:

- **Temper (temperamento)**: Variable que refleja el estado emocional del orador. Incluye tres categorías principales: depresivo, amistoso y agresivo. El valor que devuelve está dividido a su vez en dos apartados: una valoración del 1 al 100, y un nivel: bajo, medio o alto.

  El nivel alto de temperamento o *High Temper* se produce cuando el orador experimenta y expresa emociones agresivas, orientadas hacia fuera, tales como resistencia, enfado, odio, hostilidad, agresividad y/o arrogancia.

  El nivel medio de temperamento o *Medium Temper* tiene lugar cuando el hablante experimenta emociones positivas, comunicadas de una manera cálida y amistosa, tales como empatía, aceptación, amistosidad, cercanía, amabilidad, afecto, amor, calma y/o motivación. También se puede dar este nivel cuando el orador está en un estado de autocontrol de sus emociones y expresa cierta neutralidad.

  El nivel bajo de temperamento o *Low Temper* ocurre cuando el orador experimenta emociones negativas o de naturaleza inhibidora, tales

Figura 3.21: BeyondVerbal: Valores de Valence

Figura 3.22: BeyondVerbal: Valores de Arousal

como tristeza, dolor, sufrimiento, inferioridad, autoculpa, autocrítica, arrepentimiento, miedo, ansiedad y preocupación (también puede ser interpretada como fatiga). Se podría describir como si el sujeto estuviera menguando, empequeñeciendo o retirándose.

- **Valence (valencia)**: Variable que se refiere al nivel de positividad o negatividad en el que se encuentra el estado emocional. Se mueve entre estado emocional negativo y estado emocional positivo. El valor devuelto se refleja como un número en una escala de 1 a 100 y también como grupo de valencia, que puede ser negativo, neutral o positivo.

  Valencia negativa o *Negative Valence* implica que el hablante transmite dolor emocional y debilidad o agresividad o emociones antagónicas.

  Valencia neutral o *Neutral Valencia* implica que el orador no transmite ninguna emoción o muestra un estado de autocontrol.

  Valencia positiva o *Positive Valence* ocurre cuando el orador transmite emociones positivas como felicidad, calidez, entusiasmo o calma.

- **Arousal (excitación)**: Esta variable mide el nivel de energía del orador con valores que van desde la tranquilidad, el aburrimiento o el adormecimiento hasta el estado de alerta o excitación. El valor de la excitación también puede corresponder a la implicación y nivel de estimulación que siente el orador. El valor que devuelve está dividido a su vez en dos apartados: una valoración del 1 al 100 y un nivel: bajo, neutral o alto.

  Una excitación baja o *Low Arousal* corresponde a niveles de bajos de alerta y se puede registrar en casos de tristeza, comodidad, alivio o adormecimiento.

  Una excitación neutral o *Neutral Arousal* se refiere a que el orador

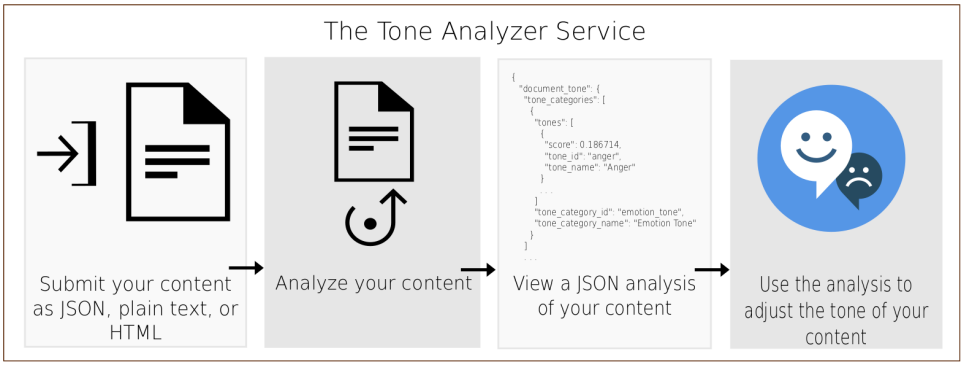

Figura 3.23: IBM Watson Tone Analyzer API: Proceso de análisis

transmite un nivel medio de alerta y se suele registrar en casos de conducta normal, indiferencia o autocontrol.

Excitación alta o *High Arousal* corresponde a altos niveles de alerta como agitación, sorpresa, comunicación apasionada, felicidad extrema o enfado.

Teniendo en cuenta estos valores y el grupo de emociones en los que se engloban, se hace una estimación en el entorno de análisis de la reacción del público ante dichas emociones y parámetros de la voz del orador.

Tone Analyzer API es una de las herramientas seleccionadas en el proyecto para el análisis del contenido del discurso. Para ello, primero se procesan los segmentos de audio del discurso del orador para convertir el audio en texto y, posteriormente, se envía este texto a la API para extraer la emoción del contenido. Se ha seleccionada esta herramienta debido a su facilidad de integración y a que dispone de una versión gratuita y de un entorno muy bien preparado para hacer pruebas.

Tone Analyzer API saca partido del análisis lingüístico cognitivo para identificar una serie de tonos en ambos niveles, de frase y de documento. Esta información se puede utilizar después para refinar y mejorar las comunicaciones. En el texto detecta tres tipos de tonos, incluidos la emoción (ira, asco, miedo, alegría y tristeza), actitudes sociales (franqueza, diligencia, extroversión, simpatía y rango emocional) y estilos de lenguaje (analítico, seguro y vacilante).

Capítulo 4

Experimentos

Tal como se ha descrito en el apartado de objetivos, una de las principales metas de este proyecto ha sido realizar varios experimentos para validar la eficacia del sistema y afinar las reacciones de la audiencia virtual.

Esta etapa de experimentación es igual de importante que la etapa de desarrollo del proyecto, dado que es el momento en el que se ha validado la efectividad de las componentes del proyecto, y ha servido para afinar las reacciones de la audiencia del entorno virtual.

4.1. Propósito del experimento

En este experimento el objetivo principal es demostrar la efectividad del juego Hola UniVRso, es decir, comprobar el grado de inmersión que sienten los oradores al estar jugando, cuáles son las reacciones que les genera el sistema y si estas reacciones son acertadas (comparándolas con reacciones de una audiencia real). En paralelo, se pretendió obtener datos sobre la efectividad general de la herramienta desarrollada: si es un método que consideran de ayuda para entrenar a hablar en público, y las partes que se podrían mejorar o cambiar.

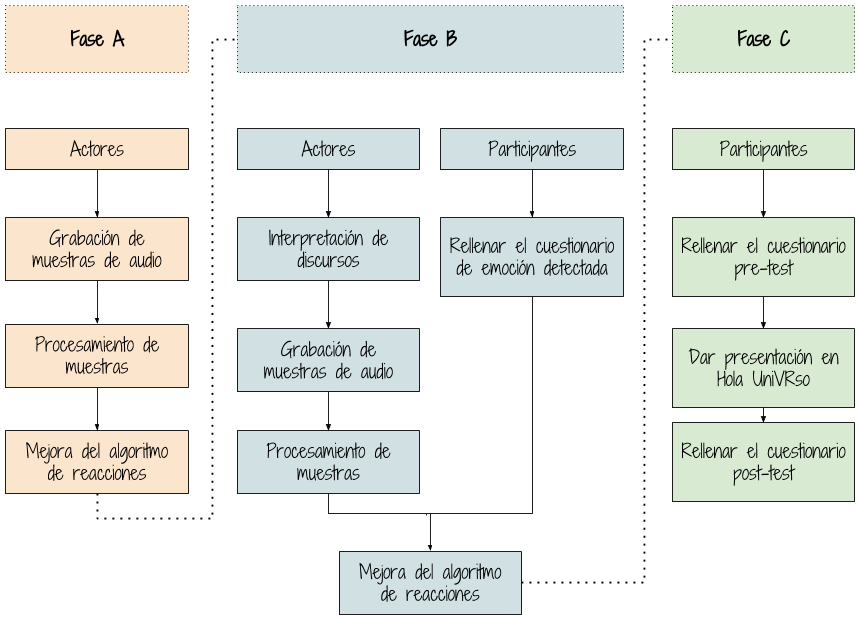

Figura 4.1: Diseño experimental. Fases del experimento

4.2. Metodología

4.2.1. Diseño experimental

El experimento se ha organizado en tres fases: fase A, fase B y fase C, resumidas en la figura 4.1.

En la **fase A** del experimento se han realizado pruebas con la librería elegida para hacer la extracción de emociones para, posteriormente, implementar el algoritmo de generación de reacciones en el entorno de análisis. Esta primera fase ha consistido en alimentar a la herramienta con fragmentos de audio grabados por dos actores interpretando los cinco grupos de emociones que reconoce la API para comprobar su efectividad y, a partir de los resultados obtenidos, poder crear un algoritmo adecuado de reacciones para la audiencia virtual.

Para esta parte solamente se ha utilizado el entorno de análisis, que ha

servido para hacer las peticiones pertinentes a la API y mostrar los resultados obtenidos. El conjunto de muestras estaba formado por 10 audios, cada uno con una duración aproximada de 30 segundos. Cada emoción está representada dos veces en el conjunto de muestras, una por un actor y otra por una actriz.

Puesto que una parte de las reacciones de la audiencia se basa en la efectividad del discurso del orador, para medir dicha efectividad se utiliza la herramienta de BeyondVerbal, una API que ofrece análisis de emociones a partir de fragmentos de audio.

Una vez implementada la primera versión del algoritmo de reacciones, se ha procedido con la segunda fase del experimento, la **fase B**. Para ello se ha contado con la ayuda de tres actores que han interpretado varios fragmentos, reflejando diversas emociones frente a un público de 20 personas. Los actores y el público contaban con un cronómetro que marcaba el tiempo de cada uno de los discursos. El audio de dichos fragmentos se ha procesado en el entorno de análisis para comprobar las reacciones generadas por el sistema. Al mismo tiempo, cada sujeto del público disponía de una hoja de encuesta de evaluador, en la que podía reflejar los cambios de emoción detectados en el discurso de cada uno de los actores y el momento temporal en el que esto ocurría.

Puesto que no todo los sujetos reflejan los cambios emocionales del discurso exactamente en el mismo segundo, se han comparado los resultados estableciendo franjas de 15 segundos, es decir, si un sujeto del público detectaba una emoción en el momento 0:16 y otro sujeto lo detectaba en el 0:26, a efectos de este experimento, se ha considerado como el mismo cambio detectado. Además de detectar el cambio de emoción y la emoción percibida en el discurso, el público también tenía la posibilidad de valorar del 1 al 10 cómo de efectivo le resultaba el discurso.

Después de analizar los resultados de la fase B, se han obtenido algunas conclusiones que han permitido afinar un poco más el algoritmo de reacciones del sistema. Por ese motivo, la última fase de este experimento, la **fase C**, ha consistido en hacer una primera prueba con el entorno completo de Hola UniVRso en un grupo de 23 personas.

Para ello, se ha organizado una jornada en la que cada uno de los participantes en el experimento entraba en el entorno virtual y daba una charla de 3 minutos. Algunos de los sujetos han preparado el tema de su charla con anterioridad y otros han improvisado en el momento. A efectos de este experimento no ha sido relevante, dado que no se ha tomado en cuenta el contenido de la charla, sino que solamente se ha analizado el audio para

generar emociones a partir de la forma del discurso y no del contenido. De forma similar a la fase anterior, a los sujetos de prueba se les ha dado un cuestionario de orador. Antes de hacer la prueba con el prototipo, los participantes debían rellenar la primera parte de este cuestionario, y después de la prueba debían rellenar el resto.

4.2.2. Materiales e instrumentos

Para la primera fase se ha utilizado un micrófono para grabar una serie de 10 audios de 30 segundos con los que posteriormente se alimenta a la API de BeyondVerbal, una herramienta de análisis de emociones seleccionada para extraer las emociones de la voz del discurso de los oradores.

En la segunda fase se ha utilizado únicamente la parte de análisis emocional o entorno de análisis de este trabajo, que consiste en una API a la que se pueden enviar una serie de medidas del orador (tono de voz y hacia dónde mira) y, a partir de esas medidas, se obtienen una serie de reacciones que muestran la efectividad del discurso del orador reflejada en forma de porcentaje. Para esta fase, como en la fase anterior, se ha usado también un micrófono y una cámara para grabar 4 muestras de audio interpretadas por 3 actores, cuya duración oscila entre los 4 y los 7 minutos cada una.

En la fase C se ha utilizado el sistema Hola UniVRso al completo.

Tanto en esta sección como en la sección de resultados (4.3) se han utilizado las siguientes abreviaturas para calificar los grupos de emociones:

A: Aburrido / Inseguro / Triste
C: Calmado
E: Enfadado / Agresivo
F: Feliz / Entusiasmado / Amistoso
N: Neutro

Cuestionarios

Tanto en la fase B como en la fase C los participantes han tenido que rellenar unos cuestionarios, los cuales han servido para generar los datos.

- **Cuestionario de la fase B**. En este cuestionario el objetivo era conseguir datos sobre cómo afectaba el discurso de los actores a los sujetos,

Variable medida	Tipo	Cómo se calcula	Rango
Segundo de cambio	Tiempo en MM:SS	Se utilizan franjas de 10 segundos para detectar cambios únicos	[00:00 - 07:00]
Engagement (¿Conectado?)	Sí o No	Cada Sí suma 1 al resultado	[0 - 1]
Engagement (Grado)	Selección múltiple	Cada selección contribuye 1 al total	[1 - 5]
Emoción	Selección múltiple	Moda de todos los valores	A, C, E, F, N

Tabla 4.1: Cuestionario de la fase B

Variable medida	Tipo	Cómo se calcula	Rango
Autovaloración del orador	Escala de 10 puntos	Media de los valores	[1 - 10]
Factores para un buen discurso	Valores del 1 al 5	Suma de los valores (el menor es el más importante)	[1 - 5]
Factor que más miedo da al hablar en público	Selección múltiple	Cada selección contribuye 1 al total	[1 - 5]
Factor donde nota el público los nervios	Varias opciones	Cada selección contribuye 1 al total	[1 - 5]
Un sistema de VR puede mejorar la oratoria	Sí o No	Suma de los valores de cada opción	[0 - 1]

Tabla 4.2: Cuestionario de la fase C

para posteriormente hacer una comparación con los resultados obtenidos por la API. Para ello, se ha generado un cuestionario que mide tres valores principales: el segundo en el que se produce un cambio de emoción en el discurso, el *engagement* o cómo de conectado está el sujeto con el discurso y la emoción detectada. Este cuestionario seguía el esquema descrito en la figura 4.1.

- **Cuestionarios de la fase C**. Para esta fase se han hecho dos cuestionarios, uno pre-test, el cual consiste en una serie de preguntas de selección múltiple de valoraciones genéricas sobre el asunto de hablar en público por parte de los participantes en la prueba, y otro post-test, más enfocado a la herramienta, para conocer en detalle cómo ha sido la experiencia de cada sujeto, qué cosas le han parecido efectivas y qué cosas cambiaría. El cuestionario pre-test tenía la forma descrita en la figura 4.2.

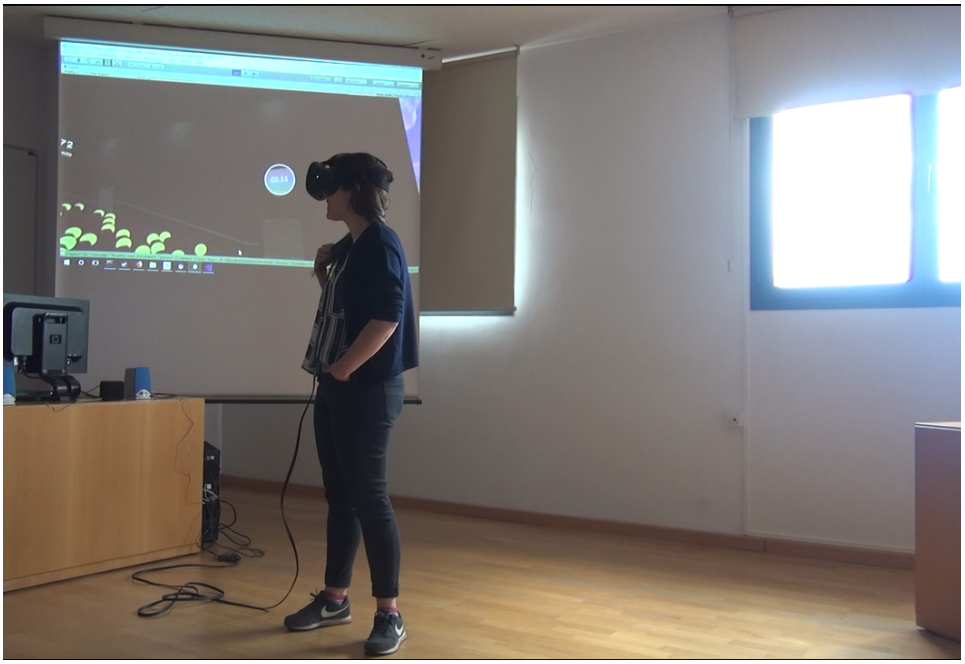

Figura 4.2: Experimento. Fase C. Sujeto probando el sistema HOLA UNIVRSO

4.2.3. Participantes

Los participantes de la primera fase han sido $n = 2$ actores, un hombre y una mujer. Los participantes de la segunda fase han sido en total $n = 19$, de los cuales 3 son actores que se han encargado de grabar las muestras de audio, y el resto son alumnos o profesores de la Facultad de Informática de la Universidad Complutense . Así mismo, en la fase final, el número de participantes ha sido un total de $n = 23$, todos ellos también alumnos o profesores de la Facultad de Informática, de entre los cuales el 52 % considera que sus habilidades para hablar en público son bajas (menor que 5 en una escala del 1 al 10). No se ha tenido que eliminar a ninguno de los sujetos participantes en este experimento puesto que no ha habido problemas a lo hora de participar en las distintas fases. En las figuras 4.2, 4.3 y 4.4 se puede ver a algunos sujetos realizando la prueba en la fase C.

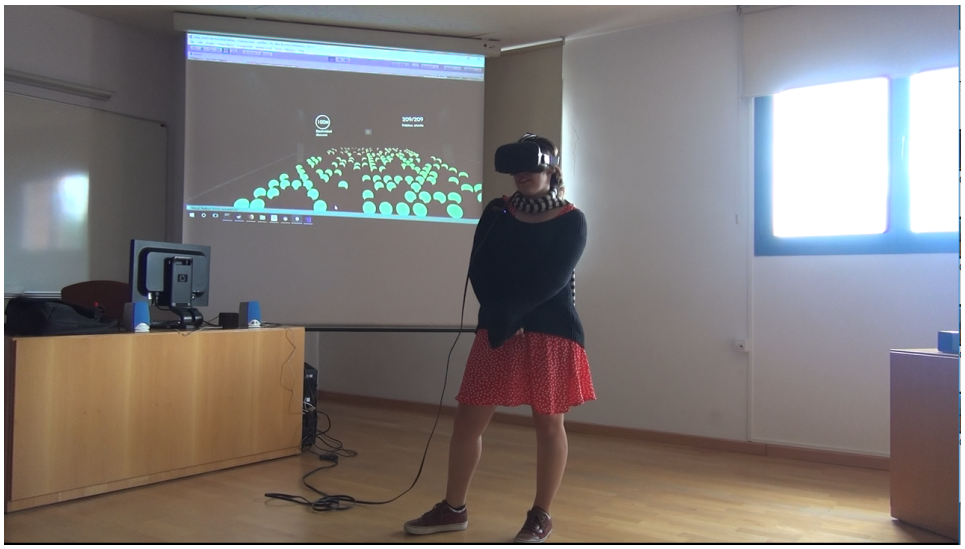

Figura 4.3: Experimento. Fase C. Sujeto probando el sistema Hola UniVRso

Figura 4.4: Experimento. Fase C. Sujeto probando el sistema Hola UniVRso

Muestra	Emoción representada	Grupo obtenido	Acierto
1.wav	Enfado	E	Sí
2.wav	Enfado	E	Sí
3.wav	Aburrimiento	A	Sí
4.wav	Aburrimiento	A	Sí
5.wav	Calma	A	No
6.wav	Calma	A	No
7.wav	Felicidad	E	Sí
8.wav	Felicidad	E	No
9.wav	Neutro	N	Sí
10.wav	Neutro	A	No

Tabla 4.3: Experimento: Resultados de la fase A

Segundo	*Engagement* ¿Conectado?	*Engagement* Grado	Grupo (AR)	Grupo (AV)
0:15	N/A	N/A	C	E
0:30	N/A	6,75	F	C
0:45	N/A	N/A	F	F
1:00	N/A	N/A	F	F
1:15	N/A	7,33	A	N
1:30	N/A	7,33	N	N
1:45	N/A	8,75	E	E
2:00	N/A	6,75	N/A	E
2:15	N/A	7,4	A	N
2:30	N/A	N/A	A	E
2:45	N/A	N/A	F	E
3:00	N/A	7	F	F

Tabla 4.4: Experimento: Resultados de la fase B. Discurso 1

4.3. Resultados

En se muestran los resultados obtenidos en todas las fases del experimento. En la fase A se han obtenido los resultados mostrados en la tabla 4.3.

En la fase B se han obtenido los resultados mostrados en las tablas 4.4, 4.5, 4.6 y 4.7[1].

En la fase C se han obtenido los resultados mostrados en la tabla 4.8.

En esta última fase, y a partir de los resultados, se han generado unos gráficos aclarativos de los datos agregados obtenidos a partir de los cuestionarios de esta fase. Estos gráficos se pueden ver en las figuras 4.5 , 4.6 y 4.7.

[1] AR: Audiencia real, AV: Audiencia virtual

Segundo	***Engagement*** **¿Conectado?**	***Engagement*** **Grado**	**Grupo (AR)**	**Grupo (AV)**
0:15	Sí	8,75	E	E
0:30	Sí	4	A	E
0:45	N/A	N/A	N/A	N/A
1:00	N/A	6	C	N/A
1:15	N	N/A	N	E
1:30	N/A	5	N/A	N/A
1:45	N/A	7,5	N/A	N/A
2:00	Sí	7,5	F	E
2:15	N/A	5	F	E
2:30	N/A	N/A	N/A	E
2:45	Sí	7,5	F	E
3:00	N	N/A	E	E
3:15	Sí	7	E	E
3:30	N	5,8	A	E
3:45	N	7,66	A	C
4:00	N	5	N/A	N/A
4:15	N	N/A	N/A	N/A
4:30	N	4	A	C
4:45	N/A	N/A	N/A	C
5:00	Sí	7	A	E
5:15	Sí	5,57	A	C
5:30	N/A	5	N/A	N/A
5:45	N/A	4	A	E
6:00	N	5	N	N
6:15	N	2,5	A	N
6:30	N	9	A	N

Tabla 4.5: Experimento: Resultados de la fase B. Discurso 2

Segundo	*Engagement* ¿Conectado?	*Engagement* Grado	Grupo (AR)	Grupo (AV)
0:15	Sí	7,6	A	E
0:30	Sí	N/A	A	E
0:45	Sí	6,5	C	E
1:00	Sí	7,14	F	F
1:15	Sí	4	C	N/A
1:30	N/A	7,33	F	N
1:45	N	3	N	N
2:00	N	6,75	A	N
2:15	N	5	A	E
2:30	N/A	3,33	A	E
2:45	Sí	7,5	E	E
3:00	N/A	8,66	E	E
3:15	Sí	8	E	F
3:30	Sí	7,71	A	A
3:45	Sí	N/A	A	A
4:00	N/A	9	N/A	N/A
4:15	Sí	7,4	C	N
4:30	Sí	8	C	N
4:45	Sí	5,66	C	N
5:00	N/A	8,8	C	C
5:15	Sí	5,57	A	A

Tabla 4.6: Experimento: Resultados de la fase B. Discurso 3

Figura 4.5: Experimento. Fase C. Factores que influyen a la hora de realizar un buen discurso

Segundo	***Engagement*** **¿Conectado?**	***Engagement*** **Grado**	**Grupo (AR)**	**Grupo (AV)**
0:15	N/A	6,50	A	A
0:30	Sí	6,50	N	N
0:45	N/A	N/A	N	N/A
1:00	N	4,50	A	C
1:15	N	3,50	A	A
1:30	N/A	4,50	N/A	N/A
1:45	N/A	8,00	N/A	N/A
2:00	Sí	6,67	N	N
2:15	N	3,33	N	E
2:30	N	N/A	N	E
2:45	N	5,33	A	N
3:00	N/A	3,50	N	E
3:15	N/A	3,00	N/A	N/A
3:30	Sí	9,00	F	C
3:45	N/A	N/A	F	N
4:00	N/A	N/A	F	E
4:15	Sí	7,17	F	E
4:30	Sí	N/A	F	E
4:45	Sí	7,50	N/A	N/A
5:00	N/A	N/A	N/A	N/A
5:15	N/A	N/A	N/A	N/A
5:30	Sí	8,00	F	E
5:45	N/A	N/A	N/A	N/A
6:00	Sí	7,00	C	C
6:15	Sí	7,50	E	E
6:30	Sí	8,50	N/A	E
6:45	N/A	6,00	E	E

Tabla 4.7: Experimento: Resultados de la fase B. Discurso 4

Valora tus habilidades como orador del 1 al 10	5,15	
Factores influyentes en una buena presentación	*Tono y proyección de voz*	3,13
	Postura	1,19
	Gesticulación	1,33
	Hacia dónde mira el orador	1,28
	Contenido de la charla	1,22
¿Qué es lo que te da más miedo al hablar en público?	*El público*	4
	Pánico a quedarte en blanco	6
	Hacer el ridículo	5
	No ser capaz de comunicar lo que quieres	11
	Que te juzguen	4
¿Dónde crees que el público nota tus nervios?	*Voz*	17
	Temblor corporal	18
	Sudor	2
	Rubor	1
	Ritmo	11
¿Crees que un sistema de realidad virtual podría ayudarte a mejorar tus habilidades como orador?	*Sí*	23
	No	0

Tabla 4.8: Experimento: Resultados de la fase C

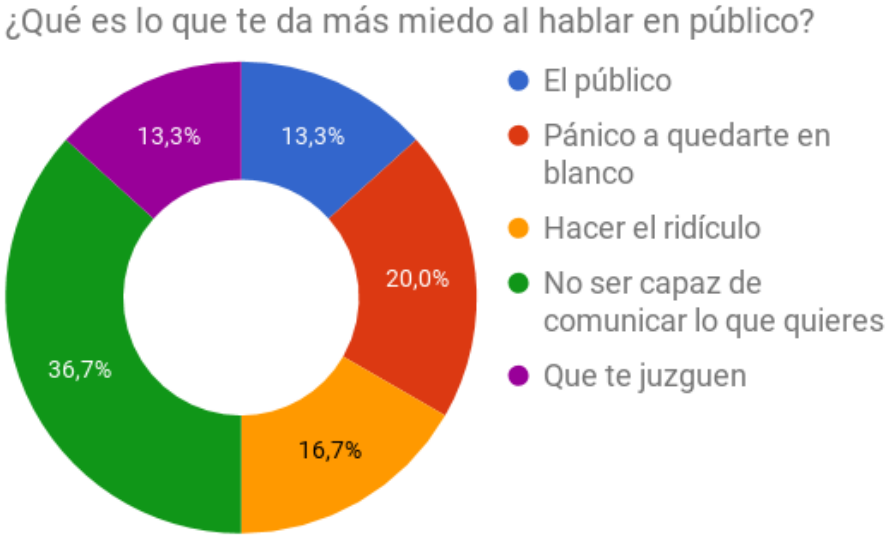

Figura 4.6: Experimento. Fase C. ¿Qué es lo que te da más miedo al hablar en público?

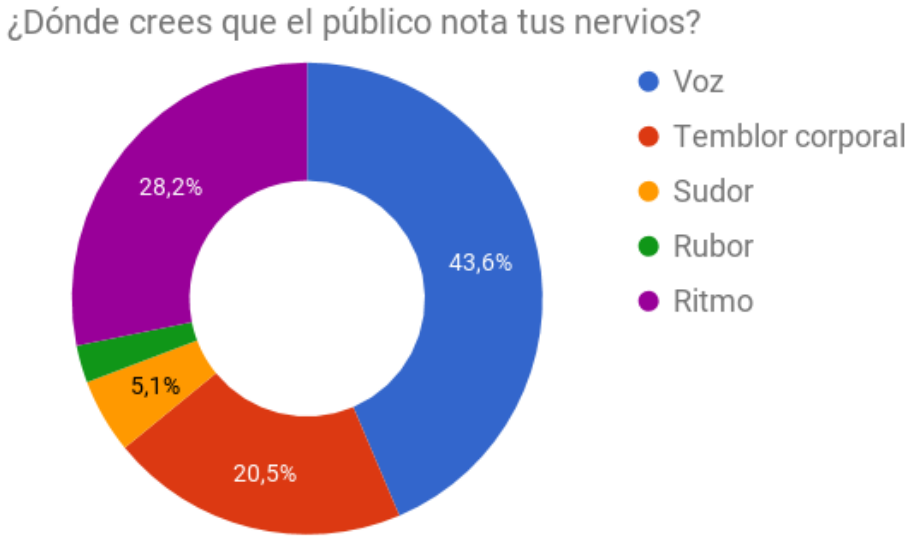

Figura 4.7: Experimento. Fase C. ¿Dónde crees que el público nota tus nervios?

4.4. Discusión

A pesar de haber obtenido en la fase A resultados acertados sólo en el 60 % de los casos (ver tabla 4.3), los resultados han permitido entender el funcionamiento de la API y han permitido asignar pesos a las distintas emociones en la generación del algoritmo de reacciones. Por ejemplo, en base a ésta y posteriores fases, se ha detectado que, en muchas ocasiones, la herramienta clasifica de forma errónea discursos entusiastas con el grupo de emociones Enfado/Disgusto/Estrés. Esto puede deberse a que el sistema de predicción que utilizan en BeyondVerbal está muy entrenado con idiomas en los que el entusiasmo se refleja de forma diferente a cómo se hace en castellano. Es más, estamos seguros de que, aunque el entrenamiento en castellano es un punto débil de la herramienta, las diferentes nacionalidades o regiones expresan las emociones de una forma diferentes.

Independientemente del motivo por el que esto ocurre, saberlo ha permitido implementar un algoritmo de reacciones teniendo en cuenta este comportamiento y otros similares.

Los resultados que se han obtenido en la fase B (ver tablas 4.4, 4.5, 4.6 y 4.7) han demostrado que el sistema de análisis de voz no tiene un alto porcentaje de efectividad en cuanto a emociones concretas, pero sí que predice con alto grado de acierto grupos similares a la emoción transmitida en cada caso. Esta fase ha permitido refinar el algoritmo en base al comportamiento del análisis de voz. También se ha constatado que una audiencia real reacciona de forma muy similar a un determinado discurso, ya que ha habido muy pocas variaciones entre los sujetos en cuanto a la emoción percibida.

Los resultados obtenidos en la última fase (ver tabla 4.8) han sido muy reveladores. Por un lado, han afianzado la utilidad de este trabajo, dado que el 100 % de los sujetos que ha realizado la prueba considera que un sistema como Hola UniVRso puede ayudarles a mejorar sus habilidades como oradores. Por otro lado, las sugerencias y críticas han servido para reorientar algunas partes de la experiencia que no se entienden o no aportan, y también para plantear o darle prioridad a algunas líneas de trabajo futuro, que se detallan en el siguiente apartado.

Basándose en los resultados de la fase C, se puede concluir que la voz, el tono y el ritmo del discurso son factores esenciales para realizar una buena presentación. Esto implica que usar una herramienta para analizar concretamente parámetros de la voz ha sido una decisión acertada, ya que se puede asumir que la reacción de una audiencia real variará conforme a las emociones y la proyección de voz del orador.

Otro dato importante que reflejan los resultados de este experimento es que lo que más miedo les produce a los participantes es la incapacidad para comunicar lo que quieren a la hora de hablar en público. Parece ser, según este resultado, que uno de los objetivos futuros de este proyecto debe focalizarse en ayudar a mejorar la comunicación efectiva, ya sea mediante pistas o consejos durante la sesión de juego, o en el informe final, para ayudar al jugador a lograr comunicar lo que realmente quiere en su discurso.

Capítulo 5

Trabajo futuro

En este capítulo se analizan las posibilidades futuras que tiene el proyecto de Hola UniVRso. Este proyecto surgio para dar una solución al miedo visceral que experimentan algunas personas al enfrentarse a un público para dar una conferencia, ya sea en el mundo académico, laboral o en la vida privada. Para validar la idea, se fue exponiendo a todo aquel que quisiera escucharla de cara a conocer su opinión al respecto y ver si tenía o no sentido. Esta técnica resultó muy acertada, dado que casi todo el mundo que escuchaba la propuesta reconocía que sufría este miedo a hablar en público o tenía a alguien cercano que lo experimentaba.

Sabiendo que, por un lado, el miedo a hablar en público es una afección generalizada y que la idea de paliarla usando la realidad virtual era algo acertado, y por otro, que el marco para un Trabajo de Fin de Máster es acotado en tiempo y recursos, se decidió crear un proyecto a partir de éste para presentarlo a un concurso de ayudas para investigación científica que organiza la Fundación BBVA. De ahí la extensión, inusualmente grande de la sección de trabajo futuro en este documento. Se pretende reflejar en ella muchas de las ideas que han surgido durante este Trabajo de Fin de Máster, y que se pretenden continuar en el proyecto más ambicioso financiado por parte de la Fundación BBVA.

5.1. Proyecto ComunicArte

A este nuevo proyecto se le ha dado el nombre de "Comunicación efectiva a través de la realidad virtual y las tecnologías educativas (ComunicArte)"[1]. Su semilla es el proyecto Hola UniVRso, pero se han aplicado los recursos en los distintos niveles, tanto de personal investigador como de tiempo y costes. Este proyecto ha sido uno entre los 10 elegidos y contará con una subvención de 100.000 euros.

Ya que este Trabajo ha sido el germen para un proyecto mucho más grande, la autora de este documento, que forma parte del equipo de investigación de ComunicARTE, ha decidido resumir dicho proyecto para dar una idea de la magnitud del trabajo futuro que va a dar lugar este proyecto.

El objetivo de este proyecto es enseñar a hablar en público utilizando videojuegos en entornos de realidad virtual.

En este proyecto se parte de la premisa de que debido al avance de la tecnología y la digitalización la empleabilidad se reducirá a la mitad. La comunicación oral se convierte en una competencia imprescindible que deben adquirir los europeos para mejorar las posibilidades de encontrar trabajo y enfrentarse al gran cambio que la digitalización supone.

La idea en ComunicArte es enseñar a hablar en público mediante videojuegos en entornos de realidad virtual. Para conseguir esto, se ha creado un equipo transdisciplinar con experiencia en diferentes campos: videojuegos educativos, narratología computacional, computación afectiva, arte, psicología y realidad virtual para crear un nuevo método de aprendizaje basado en un Entorno de Aprendizaje Gamificado Inmersivo (IGLE -*Immersive Gamified Learning Environment*) que enseñará al jugador a enfrentarse a una audiencia.

El videojuego recogerá información sobre el jugador (como medidas biométricas, análisis de voz, movimientos corporales y discurso) y valorará su oratoria en base a un modelo de audiencia virtual, que permitirá al jugador recibir consejos para mejorar su forma de comunicarse.

Por lo tanto, gracias a este nuevo marco, es posible plantear muchas mejoras que se irán implementando sobre la base construida en este trabajo.

[1] `https://bit.ly/2tgwN8R`

5.2. Nuevos sensores y nuevas mediciones

Las mediciones que se están utilizando actualmente son el análisis del audio del orador y hacia dónde dirige la mirada, pero es evidente pensar en añadir nuevos sensores y nuevas mediciones para generar reacciones de la audiencia virtual más acertadas.

Por un lado, se pueden incluir nuevos **sensores biométricos** que analicen características internas del orador como, por ejemplo, el ritmo cardíaco o la respuesta galvánica de la piel [26] para determinar si el ponente se encuentra en tensión. En base a estas medidas internas se puede hacer una comparación con lo que el orador está proyectando externamente, u ofrecerle consejos durante la sesión que le ayuden a mejorar su estado.

5.2.1. Ritmo cardíaco

Para medir el ritmo cardíaco se pueden usar sensores especializados o hacer uso de elementos cotidianos como las pulseras de actividad.

5.2.2. Respuesta galvánica de la piel

Uno de los marcadores que mejor reflejan cambios emocionales es la respuesta galvánica de la piel o GSR (*Galvanic Skin Response*) [18], también llamada conductividad de la piel o actividad electrodérmica. Este marcador modula la cantidad de sudor secretada por las glándulas sudoríparas. Esta cantidad de sudor varía según las regiones del cuerpo humano, pero se conocen sus valores normales si el individuo está relajado. Por lo tanto, es fácil determinar cuándo hay exceso de sudoración y esto puede ser debido a que el individuo se encuentre en una situación de estrés.

Por otro lado, se pueden añadir **sensores externos** que complementen a los que ya se están usando. Ahora mismo el proyecto capta la mirada del orador, pero no se está valorando la postura y la gesticulación. Esto se puede conseguir agregando un sensor externo de tipo cámara que pueda analizar los movimientos del orador, la posición de los brazos y las piernas, si está gesticulando en exceso, etc. En el mercado hay diferentes opciones para poder hacer esto.

5.2.3. Kinect

Podría utilizarse el dispositivo Kinect[2], un dispositivo con cámara integrada que permite detectar los gestos de los usuarios y muestra un esqueleto del movimiento que va haciendo el jugador en cada momento.

5.2.4. Capturadora de movimientos Smartsuit Pro

Otra opción para esta tarea es utilizar un traje Smartsuit Pro[3], una capturadora de movimiento que podría usarse para ver los movimientos del orador, o incluso para modelar los movimientos de los personajes 3D que forman el público, para que éstos sean más realistas y representen reacciones que se asemejen más a reacciones humanas.

5.2.5. Proyecto LitSens

Una de las posibles sinergias con proyectos que están actualmente en desarrollo en la Facultad de Informática de la Universidad Complutense es con el proyecto LitSens [24, 23]. LitSens es un sistema de sonido adaptativo con la capacidad de producir bandas sonoras en tiempo real que se amoldan a eventos de juego con un valor emocional asociado. El siguiente reto para este sistema, aún en desarrollo, es lograr reconocer emociones en tiempo real, de manera que no sea necesario confiar en eventos previamente clasificados por un juez humano. Con este fin, actualmente se está experimentando con prototipos que hacen uso de redes neuronales para analizar movimientos del jugador en eventos con una fuerte carga emocional. El conocimiento extraído podría ser de utilidad a la hora de reconocer movimientos asociados a emociones humanas y, por tanto, generar reacciones basadas en la gesticulación o movimientos de la cabeza del orador.

[2] `https://www.xbox.com/es-ES/xbox-one/accessories/kinect`
[3] `https://www.youtube.com/watch?v=YQAd72oIa34`

5.3. Capturar el perfil de orador

Sería interesante pensar en **clasificar a los oradores según su perfil** y para ello habría que establecer una métrica. En las escuelas de teatro existen diferentes formas de clasificar a los actores. Se suelen usar, entre otros, los elementos -aire, agua, tierra y fuego-, el eneagrama -9 tipos de personalidades diferentes-, etc. Para hacer esta clasificación, se podría ampliar la escena de *Menu* del entorno virtual con una serie de preguntas antes de empezar el juego para poder clasificar al jugador. Esto permitiría adaptar el juego al tipo de orador que esté jugando en cada momento.

Una opción para definir el perfil del orador podría ser basarse en los 4 elementos para definir 4 tipos de oradores: fuego, aire, agua y tierra, cada uno con unas características muy básicas dependientes de algunas variables diferentes (e.g. ritmo, movimiento corporal y tono de voz). Una vez perfilado el orador, el sistema juzga diferente en función del perfil. Por ejemplo, un orador que ha sido clasificado como fuego (que generalmente se acelera) tendrá una recompensa si baja su ritmo habitual. Mientras que si es tierra (orador de tipo más calmado), obtendrá una recompensa por parte de la audiencia si utiliza de vez en cuando un ritmo más acelerado para hacer énfasis en alguna parte de la ponencia.

5.4. Componente social

Ahora mismo, el juego es individual y se ejecuta en su totalidad contra la máquina, es decir, el orador hace su presentación y va recibiendo realimentación por parte de la audiencia, pero no existe intervención por parte de otros en la experiencia. Sin embargo, los algoritmos de análisis de emociones no siempre son acertados y por ello hay que **incluir el concepto de evaluación** para que el entorno se ajuste más a una situación del mundo real.

Una de las formas de aprender a hablar en público es evaluar a otros. Se puede incluir en el sistema la posibilidad de que alguien que no está jugando pueda entrar en el entorno virtual o simplemente verlo con pantalla dividida (orador | público) para aprender de lo que está haciendo otro y de la reacción del público a las acciones y emociones que va transmitiendo en cada momento. En este punto, es factible pensar en incluir jugadores externos cuya tarea sea formar parte del público y emitir juicios a través de un sistema sencillo de valoración, ya que esto enriquecería las reacciones de la audiencia virtual con reacciones reales, y daría al orador una valoración más acertada de su discurso. El sistema pasaría a tener entonces dos tipos de jugadores: oradores y evaluadores.

Es importante que los evaluadores no necesiten el sistema de realidad virtual. La mecánica sería sencilla: el evaluador ve al orador en una pantalla y, en función de lo que ve, puede dar órdenes a su avatar (que el orador está viendo en el entorno virtual) acerca de cómo se tiene que comportar para que el orador tenga realimentación instantánea. Una vez terminada la exposición, el evaluador puede hacer una evaluación más detallada del orador.

5.5. Mejorar algoritmo de reacción

El algoritmo de cálculo de reacciones actual tiene en cuenta las reacciones provenientes de los distintos análisis de características del orador, y lo pone en común con diferentes pesos asignados a cada tipo de características, teniendo en cuenta los segmentos de reacciones anteriores. Podría mejorarse este algoritmo para ofrecer cambios en la audiencia más realistas, y que se ajusten más al momento exacto en el que el orador hace una determinada acción. Para ello, se pueden considerar otros factores como el tipo de errores que comete más frecuentemente el orador, el perfil del orador, el tipo de presentación que se está haciendo, etc.

5.6. Nuevos niveles

En este momento el juego cuenta con dos niveles relativamente parecidos en aspecto, aunque uno de ellos es en un entorno más pequeño que el otro. Un trabajo futuro importante en esta línea es diseñar nuevos niveles que representen diferentes situaciones de la vida real, en los que un individuo puede experimentar ese miedo a hablar en público: una entrevista de trabajo, intervenir en clase, dar un discurso en un grupo más reducido, etc. Puesto que el sistema es modular y el entorno virtual está desacoplado del entorno de análisis, estos niveles pueden desarrollarse utilizando otras tecnologías o motores gráficos y enlazarse directamente con el entorno de análisis.

5.7. Público más realista

En la línea del apartado anterior, no sólo diseñar nuevos niveles es importante, sino también mejorar los ya existentes. Un punto crucial a mejorar es el realismo del público. En apartados anteriores se ha mencionado la decisión tomada de reducir los personajes del público a simples esferas, por cuestiones de tiempo y con el propósito de validar la efectividad del prototipo, además de disponer de estudios que demuestran que lo importante en los avatares virtuales es un comportamiento coherente. Sin embargo, como trabajo futuro, sería interesante modelar unos personajes 3D más realistas que permitan al orador empatizar con su audiencia y hagan que la experiencia se asemeje al máximo a la realidad.

5.8. Subir presentaciones en PDF y otros formatos

Tal como hacen muchas de las aplicaciones evaluadas en el apartado de Estado del arte, sería interesante ofrecerle al orador la posibilidad de subir su propia presentación al sistema en formato PDF u otros formatos. Incluso se podría enlazar con alguna API, como la API que ofrece Google Docs, para poder mostrar presentaciones que el usuario tenga en línea permitiéndole navegar entre ellas y seleccionar la que más le interese. Disponer de una presentación propia asemejaría el juego más a una situación real y permitiría al usuario practicar de forma más concreta.

5.9. Comunicación a través de sockets

La comunicación entre el entorno virtual y el entorno de análisis se hace mediante peticiones `HTTP`, que es un método adecuado si la comunicación es esporádica. Sin embargo, tanto el envío de información de las características del orador como la petición de reacciones a la API son dos métodos que se hacen a intervalos regulares y de forma muy frecuente. Por eso, como trabajo futuro, se puede realizar la comunicación mediante sockets, en lugar de hacerlo con peticiones `HTTP`, ya que éstas tienen una sobrecarga mayor para el sistema y podría ralentizarse en caso de procesar mucha información. En cambio, los sockets proporcionan una vía bidireccional de comunicación de información que sólo tiene sobrecarga en el momento de crear y destruir el socket, que permanecería abierto durante toda la sesión de juego.

Capítulo 6

Conclusiones

Para concluir, en este documento se ha visto que hablar en público es una disciplina transversal a muchos aspectos de la vida del ser humano, abarcando tareas muy diversas: dar una conferencia, hablar en una reunión de vecinos o afrontar una entrevista de trabajo. Para conseguir dominar esta disciplina, el entrenamiento se ha demostrado como una de las vías más efectivas.

En el mercado existen multitud de herramientas que utilizan la tecnología, concretamente la realidad virtual, para permitir entrenar habilidades. Los pioneros en este área son los simuladores, que se usan en diferentes ámbitos (militar, salud, industria o educación) para ofrecer el entrenamiento de diversas habilidades en un entorno virtual, haciendo que sea seguro simular situaciones que en la vida real podrían no serlo, además de reducir considerablemente los costes.

Se ha visto también que el miedo a hablar en público es una afección generalizada de la población, llegando incluso al grado de fobia (glosofobia) en algunos casos. Este miedo se puede afrontar de diversas maneras: recurriendo a técnicas con profesionales, como la reestructuración cognitiva, o mediante el uso de herramientas que usan la tecnología para el tratamiento de fobias.

Por estas dos razones surge el trabajo Hola UniVRso, con la idea de aportar en los dos sentidos: ofrecer una herramienta de entrenamiento de la habilidad de hablar en público, en la que los oradores pueden practicar sus presentaciones en un entorno seguro; y servir como herramienta de ayuda a las técnicas profesionales para superar o paliar el miedo a hablar en público, sirviendo como paso previo a la exposición a la situación real.

A pesar de que existen aplicaciones que intentan paliar el miedo a hablar en público mediante realidad virtual, de forma similar a lo que se propone en ese proyecto, en ninguno de las aplicaciones investigadas para este trabajo se ha encontrado una herramienta que proporcione un público reactivo en tiempo real a parámetros del orador.

Para conseguir este objetivo, se ha construido un sistema, en forma de videojuego, separado en dos grandes bloques: entorno virtual y entorno de análisis. El entorno virtual es el lugar en el que tiene lugar el juego, se posiciona al orador frente a una audiencia virtual y ahí procede a dar su exposición durante un tiempo limitado. En ese entorno, el orador puede ir comprobando la efectividad de su discurso y adaptarlo a las reacciones de la audiencia.

Dichas reacciones se generan en el entorno de análisis, donde el tono de voz, el contenido del discurso y la mirada del orador sirven como parámetros para extraer emociones de su discurso y generar reacciones en la audiencia del entorno virtual.

El primer prototipo de Hola UniVRso se ha probado con un grupo de individuos y se han obtenido sus valoraciones de la herramienta. Los resultados de este experimento han corroborado tres puntos fundamentales. Según los participantes del experimento:

- Una herramienta de realidad virtual como la propuesta en este trabajo puede ser efectiva para practicar a hablar en público.
- Una audiencia que proporciona realimentación al orador en tiempo real es más efectiva que un informe al final de la experiencia, dado que el orador se enfrenta a una situación más parecida al mundo real, y debe modificar su discurso y adaptarse a su público en cada momento.
- Una de las componentes que más influyen a la hora de dar un buen discurso es el tono y proyección de la voz del orador.

Gracias a estos resultados, se ha validado la utilidad de la herramienta construida en este trabajo, poniendo de manifiesto la decisión acertada de utilizar un análisis de la voz del orador para extraer emociones transmitidas en en el discurso del orador, y a partir de dichas emociones generar reacciones en la audiencia virtual.

Por último queda indicar que este proyecto es el precursor y la primera versión de un proyecto de investigación más amplio que cuenta con más recursos económicos y de tiempo. La arquitectura de Hola UniVRso se ha diseñado de forma modular, por lo que resultará sencillo implementar las

mejoras propuestas en el apartado de Trabajo futuro en el marco de este nuevo proyecto.

Chapter 6

Conclusions

To conclude, it has been stated through this document that public speaking is a transversal discipline to many human life aspects, such as giving a talk, speaking in a boardroom or facing a job interview. To master this discipline, training has proven to be one of the most effective methods.

There are multiple tools in the market that use technology, more specifically, virtual reality, to train abilities. Simulators have been the pioneers in this field. They are used in diverse areas (military, health, industry or education) to offer training in certain skills in a virtual environment, making it safer to simulate real world situations which may not be safe to face in real life, and reducing costs considerably.

It has also been shown that fear of public speaking is a generalized condition that affects the population, reaching in some cases the degree of phobia (glossophobia). This fear can be dealt with in multiple ways: seeking professional treatment, such as cognitive restructuring, or using certain tools that involve technology to help overcome fears and phobias.

The work of Hola UniVRso was born due to this two previous reasons, with the goal of contributing in both directions: providing a training tool for the ability of public speaking, in which orators can practice their presentations in a safe environment; and also serve as tool to aid in professional therapy to help overcome or decrease fear of public speaking, being a starting platform before the step of facing a real world audience.

Even though there are many projects that provide the tools to overcome fear of public speaking through virtual reality, in a similar way of what this work tries to achieve, none of the researched tools provides a reactive audience that evolves and reacts in real time to the speakers parameters.

To reach this goal, this project has been built as a videogame, divided in two big blocks: virtual environment and analysis environment. The virtual environment is the places where the game takes places. The orator stands in front of a virtual audience and gives a talk during a limited amount of time. In this environment, the speaker can check its speech effectiveness and adapt it to the audience reactions.

These reactions are generated in the analysis environment, where the voice tone, the content of the speech and speaker's focus act as parameters to extract emotions from the speech and generate reactions in the virtual audience.

The first prototype of this work has been tested in a group of individuals, which have been asked to give feedback about their general experience. The results obtained in this experiment have corroborated three main points. According to the experiment participants:

- A virtual reality tool such as the one in this work can be an effective method to train the ability of public speaking.
- An audience that gives real time feedback is more effective than a final report at the end of the experience, because the speaker faces a situation that resembles best a real life situation, and has to change its speech and adapt to its audience in real time
- Voice tone and projection is one of the most important factors that affect a good speech

The previous findings have validated the usefulness of the tool built in this work, showing that it was the right choice to use the speaker's voice analysis to extract emotions conveyed in its speech, and from those emotions generate reactions in the virtual audience.

Finally, it should be mentioned that this work is the precursor and first version of a bigger research project, which has been granted more economical and time resources. The architecture of HOLA UNIVRSO has been designed in a modular scheme, which will make it easy to introduce the enhancements and functionalities stated in the Future work section in the scope of this new project.

Bibliografía

Y así, del mucho leer y del poco dormir,
se le secó el celebro de manera que vino
a perder el juicio.

Miguel de Cervantes Saavedra

[1] Alan Amory, Kevin Naicker, Jacky Vincent, and Claudia Adams. The use of computer games as an educational tool: identification of appropriate game types and game elements. *British Journal of Educational Technology*, 30(4):311–321, 1999.

[2] Kyle Banker. *MongoDB in action*. Manning Publications Co., 2011.

[3] Woodrow Barfield and Suzanne Weghorst. The sense of presence within virtual environments: A conceptual framework. *Advances in Human Factors Ergonomics*, 19:699–699, 1993.

[4] Paul Boersma, Titia Benders, and Klaas Seinhorst. Neural network models for phonology and phonetics. *Manuscript in preparation*, 2013.

[5] Yoram S. Bonneh, Yoram Levanon, Omrit Dean-Pardo, Lan Lossos, and Yael Adini. Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in human neuroscience*, 4:237, 2011.

[6] Kristina Chodorow. *MongoDB: The Definitive Guide: Powerful and Scalable Data Storage*. O'Reilly Media, Inc., 2013.

[7] Marcus Tullius Cicero, James M. May, and Jakob Wisse. *Cicero on the ideal orator*. Oxford University Press New York, 2001.

[8] Columbia University Press. Gettysburg Address. Columbia Encyclopedia, 6. ed. `https://bartelby.com/`. Consultado el 2018-06-01.

[9] Roddy Cowie and Randolph R. Cornelius. Describing the emotional states that are expressed in speech. *Speech communication*, 40(1-2):5–32, 2003.

[10] Martin Ebner and Andreas Holzinger. Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Computers & education*, 49(3):873–890, 2007.

[11] Paul Ekman and Wallace V. Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.

[12] Moataz El Ayadi, Mohamed S. Kamel, and Fakhri Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3):572–587, 2011.

[13] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 835–838. ACM, 2013.

[14] Michael Fleming, Dale Olsen, Hilary Stathes, Laura Boteler, Paul Grossberg, Judie Pfeifer, Stephanie Schiro, Jane Banning, and Susan Skochelak. Virtual reality skills training for health care professionals in alcohol screening and brief intervention. *The Journal of the American Board of Family Medicine*, 22(4):387–398, 2009.

[15] Virginia Francisco Gilmartín. Identificación automática del contenido afectivo de un texto y su papel en la presentación de información. 2009.

[16] Philippe Fuchs and Guillaume Moreau. *Le traité de la réalité virtuelle*, volume 2. Presses des MINES, 2006.

[17] Marcel Gratacós. Glosofobia: Características, diagnóstico y tratamiento. `https://www.lifeder.com/glosofobia/`. Consultado el 2018-06-01.

[18] Mandy Gruber and Penny Moore. Galvanic skin response. *The Science Teacher*, 64(9):52, 1997.

[19] M. Dean Havron and Leslie F. Butler. Evaluation of training effectiveness of the 2fh2 helicopter flight trainer research tool. *Naval Training Device Center, Port Washington, NY*, 1957.

[20] Rob High. The era of cognitive systems: An inside look at ibm watson and how it works. *IBM Corporation, Redbooks*, 2012.

[21] J. Huizinga. Homo Ludens Ils 86 (vol. 3), 2014.

[22] Gwo-Jen Hwang and Po-Han Wu. Advancements and trends in digital game-based learning research: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 43(1), 2012.

[23] Manuel López Ibáñez, Nahum Álvarez, and Federico Peinado. Litsens: An improved architecture for adaptive music using text input and sentiment analysis. In *In Proceedings of the C3GI Conference 2017*, 2017.

[24] Manuel López Ibáñez, Nahum Álvarez, and Federico Peinado. Towards an emotion-driven adaptive system for video game music. In *International Conference on Advances in Computer Entertainment*, pages 360–367. Springer, 2017.

[25] Borja Manero Iglesias. *Del teatro clásico a los videojuegos educativos.* Universidad Complutense de Madrid, 2015.

[26] iMotions. What is gsr (galvanic skin response) and how does it work? `https://imotions.com/blog/gsr/`, 2015.

[27] Martin Luther King Junior. I have a dream.

[28] Borja Manero, Javier Torrente, Clara Fernández-Vara, and Baltasar Fernández-Manjón. Investigating the impact of gaming habits, gender, and age on the effectiveness of an educational video game: An exploratory study. *IEEE Transactions on Learning Technologies*, 10(2):236–246, 2017.

[29] Borja Manero, Javier Torrente, Manuel Freire, and Baltasar Fernández-Manjón. An instrument to build a gamer clustering framework according to gaming preferences and habits. *Computers in Human Behavior*, 62:353–363, 2016.

[30] Borja Manero, Javier Torrente, Ángel Serrano, Iván Martínez-Ortiz, and Baltasar Fernández-Manjón. Can educational video games increase high school studentsínterest in theatre? *Computers & Education*, 87:182–191, 2015.

[31] Elad Maor, Jaskanwal D. Sara, Diana M. Orbelo, Lilach O. Lerman, Yoram Levanon, and Amir Lerman. Voice signal characteristics are independently associated with coronary artery disease. In *Mayo Clinic Proceedings*. Elsevier, 2018.

[32] Francisco Martínez-Sánchez, José Antonio Muela-Martínez, Pedro Cortés-Soto, Juan José García Meilán, Juan Antonio Vera Ferrándiz, Amaro Egea Caparrós, and Isabel María Pujante Valverde. Can the acoustic analysis of expressive prosody discriminate schizophrenia? *The Spanish journal of psychology*, 18, 2015.

[33] Peter Membrey, Eelco Plugge, and DUPTim Hawkins. *The definitive guide to MongoDB: the noSQL database for cloud and desktop computing.* Apress, 2011.

[34] Milana Milošević and Željko Đurović. Challenges in emotion speech recognition. In *3rd International Conference on Electrical, Electronic and Computing Engineering, IcETRAN*, 2016.

[35] James C Mundt, Peter J Snyder, Michael S Cannizzaro, Kara Chappie, and Dayna S Geralts. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (ivr) technology. *Journal of neurolinguistics*, 20(1):50–64, 2007.

[36] Janet Horowitz Murray and Janet H. Murray. *Hamlet on the holodeck: The future of narrative in cyberspace*. MIT press, 2017.

[37] Ernest H Page and Roger Smith. Introduction to military training simulation: a guide for discrete event simulationists. In *Simulation Conference Proceedings, 1998. Winter*, volume 1, pages 53–60. IEEE, 1998.

[38] Dr. Antonio Cano Vindel (Sociedad Española para el Estudio de la Ansiedad y el Estrés (SEAS)). Tratamientos eficaces. `http://webs.ucm.es/info/seas/ta/trat_efi.htm`. Consultado el 2018-06-01.

[39] Dr. Antonio Cano Vindel (Sociedad Española para el Estudio de la Ansiedad y el Estrés (SEAS)). ¿La ansiedad a hablar en público, puede llegar a ser una patología? `http://webs.ucm.es/info/seas/faq/habl_pub.htm`. Consultado el 2018-06-01.

[40] Daniel Parente. Gamificacion en la educación. *Gamificación en aulas universitarias*, 11, 2016.

[41] Sarah Parsons and Peter Mitchell. The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of intellectual disability research*, 46(5):430–443, 2002.

[42] JA. Pérez-Carrasco, C. Suarez-Mejías, B. Acha, José L. López-Guerra, and C. Serrano. Comparación de un método de segmentación de tumores retroperitoneales con herramientas comerciales de uso clínico. *LIBRO DE ACTAS*, page 47.

[43] Joseph Psotka. Immersive training systems: Virtual reality and education and training. *Instructional science*, 23(5-6):405–431, 1995.

[44] Marco Fabio QUINTILIANO. Institutio oratoria, traducción de i. *Rodríguez y P. Sandier, Madrid, Hernando*, 2:1–11, 1987.

[45] Barbara Olasov Rothbaum, Larry Hodges, Renato Alarcon, David Ready, Fran Shahar, Ken Graap, Jarrel Pair, Philip Hebert, Dave Gotz, Brian Wills, et al. Virtual reality exposure therapy for ptsd vietnam veterans: A case study. *Journal of traumatic stress*, 12(2):263–271, 1999.

[46] Maria V Sanchez-Vives and Mel Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332, 2005.

[47] Disa A. Sauter, Frank Eisner, Andrew J. Calder, and Sophie K. Scott. Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, 63(11):2251–2272, 2010.

[48] Edward Schiappa. *Protagoras and logos: A study in Greek philosophy and rhetoric*. University of South Carolina Press, 2013.

[49] Thomas B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1):120–126, 1992.

[50] Mel Slater, David-Paul Pertaub, Chris Barker, and David M. Clark. An experimental study on fear of public speaking using a virtual environment. *CyberPsychology & Behavior*, 9(5):627–633, 2006.

[51] Kaveri Subrahmanyam and Patricia M. Greenfield. Effect of video game practice on spatial skills in girls and boys. *Journal of applied developmental psychology*, 15(1):13–32, 1994.

[52] Angela Swenson. You make my heart beat faster: A quantitative study of the relationship between instructor immediacy, classroom community, and public speaking anxiety. *Journal of undergraduate research*, 14:1–12, 2011.

[53] PhD. Mattew Tull. How virtual reality exposure therapy (VRET) treats PTSD. `https://www.verywellmind.com/virtual-reality-exposure-therapy-vret-2797340`, 2018. Consultado el 2018-06-01.

[54] Dimitrios Ververidis and Constantine Kotropoulos. Emotional speech recognition: Resources, features, and methods. *Speech communication*, 48(9):1162–1181, 2006.

[55] Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. Building characters: Lessons drawn from virtual environments. In *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, pages 119–126, 2005.

[56] Dr. Antonio Cano Vindel. El pánico escénico puede durar toda la vida si no se trata con ayuda de psicólogos, 2015.

[57] Thurid Vogt, Elisabeth André, and Nikolaus Bee. Emovoice-a framework for online recognition of emotions from voice. In *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*, pages 188–199. Springer, 2008.

[58] Johannes Wagner, Florian Lingenfelser, and Elisabeth André. The social signal interpretation framework (ssi) for real time signal processing and recognition. In *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[59] Guang-Xin WANG and LI Li. Virtual reality exposure therapy of anxiety disorders. *Advances in Psychological Science*, 20(8):1277–1286, 2012.

[60] Eduard Zeller, J. Rovira Armengol, et al. Sócrates y los sofistas. Technical report, Nova, 1955.

–Andorin, ¿crees en los dioses?

–¿En qué?

–En los dioses.

–No había oído nunca esa palabra ¿Qué son?

–No es una palabra del idioma galáctico–dijo Namarti

Hacia la Fundación

Isaac Asimov